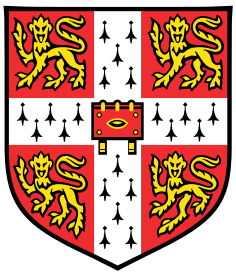
Identification of RBPMS as a smooth muscle master splicing regulator via association of its gene with super-enhancers



Erick Eidy Nakagaki Silva

Supervisor: Prof. C.W.J. Smith

Department of Biochemistry
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Churchill College

April 2019

I would like to dedicate this thesis to my parents, my sister and to the memory of my grandfather Vô Zito.

"There are three [profanity] loops!"

– attributed to Phillip Sharp
upon the discovery of RNA splicing in 1977

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Erick Eidy Nakagaki Silva
April 2019

Acknowledgements

I would like to thank those that contributed to the success of this work and made the completion of this Ph.D study possible with their kind support.

My Supervisor

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Chris Smith for being such an inspiring mentor, continuously supporting my Ph.D in the last four years.

My RNA labmates

Clare Gooding and Miriam Llorian, for the guidance in the beginning of my Ph.D as well as for their direct contribution to the completion of this work, by establishing RBPMS-A overexpression cell lines and carrying the bioinformatic analysis respectively.

The students I have had the honor to supervise during my Ph.D, George Cameron, Fredderick Richards and Zoe Heckhausen, who definitely contributed to this work with exciting results.

I also would like to thank current and former members of the RNA labs, especially Aishwarya Jacob, Angela Rubio, Anna Git, Caia Duncan, Elisa Casanova and Miguel Coelho who over the years became more than just lab colleagues.

Carrington's lab

Members of the Carrington's Lab, including Mark Carrington, Olivia Macleod, Isobel Hambleton and Camilla Trevor for the help with the equipment and the happy hours after long days in the lab.

Funding Agencies

I am also very grateful to the funding agencies, especially the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil) for the Science without Borders scholarship.

My Former Supervisors

I also would like to thank my former supervisors for their motivating supervision without which I would not have pursued this Ph.D: Dr. Marina Mourão from Centro de Pesquisa Rene Rachou (CPqRR) and Dr. Daniel Ungar from the University of York.

My Friends

My friends in Cambridge, especially the JERDIL, Janaina Nascimento, Rodrigo Pontes, Denise Dalbosco, Isabela Navarro and Luiza Peruffo, for making the best of all the moments that we were together.

I am also very grateful for having the opportunity to share my day to day life with Bruno, Giuseppe, Lavinia and Aline who together made our house our home.
My friends in Brazil, especially the ones from school, Brenda, Carolina, Fernanda, Gabriela, Heloyza, Luisa and Mariana; from university, Ana Carolina, Ana Catarina, Laura Furtado and Laura Soares; and Bruna de Queiroz from my year abroad at the University of York. All these friends supported me even at a distance.

Besides all of these people, I would like to thank my best friend Tim who has always been by my side.

My Family

Last but not the least, I would like to thank my family, especially my parents and my sister, for the constant motivation and endless support when decisions were not easy to be made.

Abstract

IDENTIFICATION OF RBPMS AS A SMOOTH MUSCLE MASTER SPLICING REGULATOR VIA ASSOCIATION OF ITS GENE WITH SUPER-ENHANCERS

Erick Eidy Nakagaki Silva

Alternative splicing (AS) is primarily regulated by regulatory RNA-binding proteins (RBPs). It has been suggested that a small number of master splicing regulators might control cell-specific programs and these regulators could be identified via the association of their genes with transcriptional super-enhancers. Using this approach, RNA Binding Protein with Multiple Splicing (RBPMS) was identified as a critical splicing regulator in vascular smooth-muscle cells (SMCs). RBPMS is strongly downregulated during SMC dedifferentiation and is responsible for nearly 20% of the AS changes during this transition as indicated by mRNA-Seq of rat PAC1 cells with RBPMS-manipulated levels. RBPMS overexpression also promoted splicing events that are usually only observed in tissue SMCs. RBPMS targeted a network of proteins involved in the cytoskeleton and cell-adhesions, machineries remodeled during the transition from contractile to motile-dedifferentiated SMCs. RBPMS directly regulated target exons with a positional bias depending upon whether acting as an activator or repressor, as indicated by RBPMS-maps, *in vivo* transfections with minigene reporters, RBPMS RNA binding mutant, MS2 artificial tethering and lastly *in vitro* binding assays. RBPMS controlled splicing and activity of other splicing and post-transcriptional regulators (MBNL1, MBNL2 and LSM14B) as well as the key SMC transcription factor *Myocardin*. Structure-function analyses revealed that the two major RBPMS isoforms (RBPMS-A and B) have differential activity, and that dimerization and RBPMS C-terminus are essential to RBPMS splicing activity. Yet, RBPMS RRM was insufficient for splicing. In fact, a core section of the C-terminus of RBPMS-B antagonized its repressor-splicing activity. Additionally, two threonine residues of RBPMS could be phosphorylated differentially modulating RBPMS isoforms activity. Therefore, this study provides the strongest evidence to date for a molecular function of RBPMS as a splicing-regulator, matching many of the expected criteria of a master regulator of AS in differentiated VSMCs.

Table of contents

List of figures

List of tables

Nomenclature

ΔPSI	Difference of Percentage of Spliced In
A	Alanine
A3SS	Alternative 3′ splice site
A5SS	Alternative 5′ splice site
ACTA2	Smooth muscle Actin
ACTN1	Actinin1
AS	Alternative splicing
ASE	Alternative splicing event
ASEID	Alternative splicing event identifier
ASO	Antisense oligonucleotide
Ad2	Adenovirus 2
Arg	Arginine
Asp	Aspartic acid
BSA	Bovine serum albumin
CNN	Calponin
CTD	Carboxy terminal domain
Chr	Chromosome
Cryo-EM	Cryogenic electron microscopy

D	Aspartic acid
D	Differentiated SMC
D Ctr	Differentiated PAC1 knockdown control
D KD	Differentiated PAC1 RBPMS knockdown
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
Dox	Doxycycline
E	Glutamic acid
ECM	Extracellular matrix
EMSA	Electrophoretic mobility shift assay
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
FBS	Fetal bovine serum
FDR	False discovery rate
FN1	Fibronectin
Flnb	Filamin B
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	Green fluorescent protein
GO	Gene ontology
GTEx	Genotype-Tissue Expression
Glu	Glutamic acid
H1	Hinge 1
HEK293	Human embryonic kidney 293 cells
HERMES	HEart RRM Expressed Sequence

HSA	Hot spot area
His	Histidine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
K	Lysine
KD	RBPMS knockdown
KD1	RBPMS knockdown with siRNA1
KD2	RBPMS knockdown with siRNA2
LASR	Large assembly of splicing regulators
Lys	Lysine
MBNL	Muscleblind-like protein
MRTF	Myocardin-related transcription factors
MXE	Mutually exclusive exon
MYOCD	Myocardin
Msl2	Male-specific lethal 2
N	Asparagine
NGS	Next Generation Sequencing
NLS	Nuclear localization signal
NM	Non-smooth muscle
NMD	Nonsense-mediated mRNA decay
OE	RBPMS-A overexpression
P	Proliferative\Dedifferentiated SMC
P	Proline

P Ctr	Proliferative PAC1 RBPMS-A minus dox control
P LV	Proliferative PAC1 lentiviral control
P OE	Proliferative PAC1 RBPMS-A overexpression (plus dox)
P0	Passage 0
P9	Passage 9
PAC1	Rat pulmonary artery cell line
PAP	Poly-A binding
PAR-CLIP	Photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation
PBS	Phosphate-buffered saline
PC1	Principal Component 1
PC2	Principal Component 2
PCA	Principal Component Analysis
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PPI	Protein-protein interaction
PSI	Percentage of Spliced In
PT	Post-transcriptional
PTBP1	Polypyrimidine tract binding protein
PTC	Premature termination codon
PTM	Post-translational modification
Phe	Phenylalanine
Pol II	RNA polymerase II
QKI	Quaking protein

R	Arginine
RBPMS	RNA Binding Protein with Multiple Splicing
RBPMS-A	RBPMS isoform A
RBPMS-B	RBPMS isoform B
RGC	Retinal ganglion cell
RI	Retained intron
RNA	Ribonucleic acid
RNAseq	RNA sequencing
RRM	RNA Recognition Motif
RT-PCR	Reverse transcription PCR
S	Serine
SC	Enzymatically dispersed single cultured SMCs
SDS	Sodium dodecyl sulfate
SE	Skipped exon
SF	Splicing factor
SM	Smooth muscle
SMC	Smooth muscle cell
SR proteins	Serine Arginine rich proteins
SRF	Serum response factor
Sd	Standard deviation
Sxl	Sex Lethal
T	Aorta tissue
T	Threonine
TF	Transcription factor

TPM	Transcripts per million
TPM1	Tropomyosin1
Tra	Transformer
UTP	Uridine-5′-triphosphate
UTR	Unstranslated region
UV	Ultraviolet light
VSMC	Vascular smooth muscle cells
cDNA	Complementary DNA
eCLIP	Enhanced CLIP
hESC	Human embryonic stem cells
hnRNP	Heterogeneous nuclear ribonucleoprotein
iCLIP	Individual-nucleotide resolution Cross-Linking and Immuno-Precipitation
mCAC	CAC to CCC mutant
mCAC1	Cluster 1 CAC mutation
mCAC2	Cluster 2 CAC mutation
mRNA	Messenger RNA
miRNA	microRNA
n	Sample size
ncRNA	non-coding RNA
p	p-value
p-adj	Adjusted p-value
pInd	pINDUCER vector
qRT-PCR	Quantitative PCR

siRNA	Silencer RNA
snRNPs	Small nuclear ribonucleoproteins
wt	wild-type

Chapter 1

Introduction

1.1 Gene expression control in eukaryotes

"I have also tried to relate the problem to the other central problems of molecular biology - those of gene action and nucleic acid synthesis"

– Francis Crick, On protein synthesis, 1957

The quote above was stated by Francis Crick in 1957 when addressing the protein synthesis problem which led to the postulation of the Central Dogma of molecular biology (Crick, 1958). The dogma remarkably proposed the flow of genetic information within a cell to be from DNA to RNA to finally protein (Crick, 1970). In other words, DNA is able to auto-replicate and to be decoded into RNA during transcription which then can be translated into a polypeptide sequence. Even though a few years later, multi-directionality of the flow by reverse information of RNA to DNA in virus was reported, the initially proposed transfer of genetic information, perhaps rather simple, still stands at the center of gene expression as the principal route.

During the genetic information flow, RNA molecules intermediate the translation of the genetic code into proteins, yet their synthesis is coordinated by the signals embedded in the DNA template. The *cis*-elements in the DNA mediate recruitment of *trans*-acting components that comprise a plethora of transcription factors. The *cis*-elements include enhancers, promoters, transcription terminators and others. These signals play a major role in the recruitment and release of the RNA polymerase II (pol II), the main enzyme of the transcription apparatus. Pol II catalyzes the production of protein coding mRNAs and a number of non-coding RNAs (ncRNA) in eukaryotic cells. In addition, chromatin remodeling factors play a critical role in the control of

transcription by regulating the physical accessibility to the regulatory sequences found in the DNA. Thus, specific cell states can be produced by the establishment of gene expression programs through control of the first step in decoding the DNA sequence by transcription factors, cofactors, and chromatin regulators. In that way, misregulation of this process is associated with a broad range of disease (reviewed in Lee and Young (2013)).

In addition to transcriptional control, gene expression can also be regulated at the post-transcriptional and translational level. Moreover, even the final protein is subjected to post-translational control. This regulates the protein's turn-over and activity by deactivation, activation or modifications. Together these processes determine the final array of proteins expressed in a cell-type.

1.2 Post-transcriptional processes as regulators of gene expression

In prokaryotes, little or no mRNA processing is observed, with ribosomes being able to assemble on the transcript and promote translation before completion of transcription. This is mainly due to the fact that these processes do not take place in separated compartments within the prokaryotic cells and also because bacterial genes are continuous, encoding genetic messages for functional proteins in an uninterrupted manner. On the other hand, in most eukaryotic cells, prior to cytoplasm export for translation, pre-mRNAs in the nucleus undergo maturation which consists of a three-step process: $5'$ capping, splicing, and $3'$ cleavage together with polyadenylation (Figure 1.2). Briefly, this involves the attachment of a 7-methylguanosine cap to the $5'$ end, after which intronic sequences are removed and exons ligated upon splicing. Finally, a poly(A) tail is added after the $3'$ end cleavage. Although commonly called post-transcriptional, the pre-mRNA processing occurs at the gene in concert with transcription. Thereby, coupling of mRNA processing with transcription has been suggested to exist with the Pol II carboxy-terminal domain (CTD) playing a critical role in the functional connection of these processes (Bentley, 2014).

Therefore, post-transcriptional processing provides another layer of gene expression regulation in eukaryotic cells. In the post-transcriptional gene control, RNA-binding proteins (RBPs) and ribonucleoproteins are the key players coordinating a range of processes including maturation, transport and stability of messenger as well as non-coding RNAs (mRNA and ncRNA) (Gerstberger et al., 2014). So, although transcriptional programs have long been studied as the main driving forces of cell

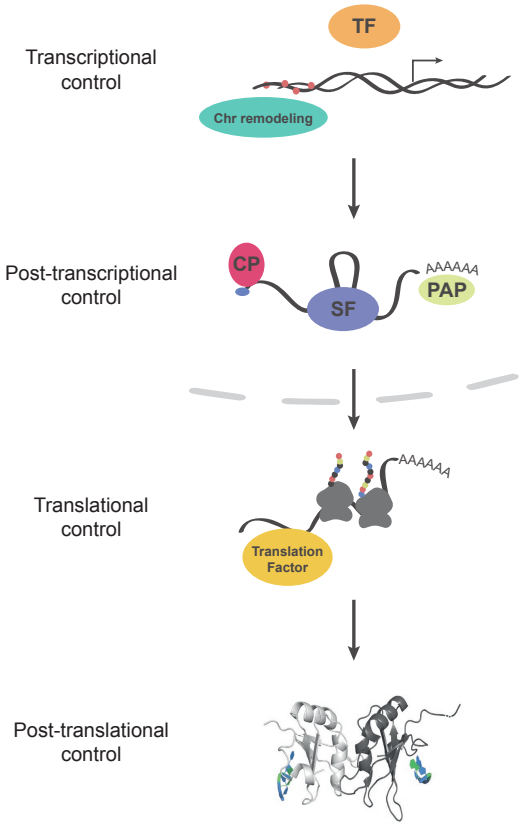


Fig. 1.1 Layers of gene expression regulation. Gene expression can be controlled at different levels. Transcriptional control regulates the transcription of the genes. Post-transcriptional control acts on pre-mRNA processing, transport and decay. Translational control determines the translation of mRNAs. Post-translational control involves the regulation of the proteins by for example activation and degradation. (TF) transcription factor, Chromatin (Chr) remodeling proteins, (CP) Cap binding proteins, (SF) Splicing factor, (PAP) Poly-A binding proteins. 5′ cap is shown in blue.

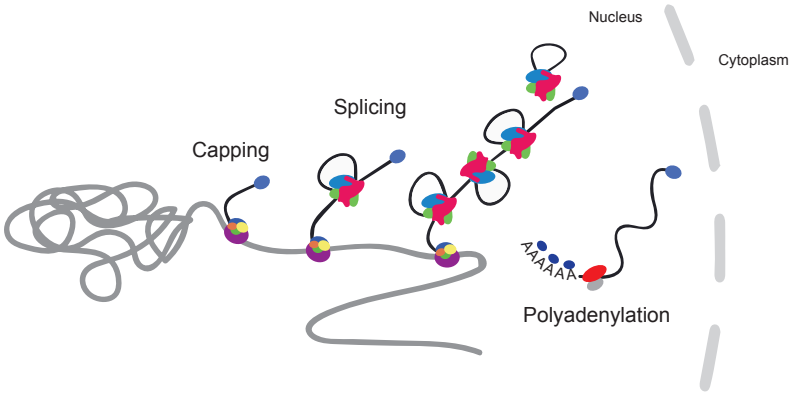


Fig. 1.2 pre-mRNA processing. pre-mRNAs while transcribed in the nucleus undergo three main modifications 5′ end capping, splicing and 3′ cleavage and polyadenylation to produce mature mRNAs. These are then exported to the cytoplasm. DNA is represented by gray lines and RNA by black lines. RNA-polymerase II and associated factors forming the transcription elongation complex is shown in dark purple, yellow, green and orange. Spliceosome is highlighted in pink, blue and green. 5′ cap is represented in light purple. Cleavage and polyadenylation is represented by red and gray. RBPs bound to the poly-A tail are shown in dark blue.

specific transcriptomes, recent lines of evidence have brought post-transcriptional regulation at the forefront of the processes contributing to the reshaping of tissue transcriptomes.

1.3 pre-mRNA splicing

1.3.1 Chemistry of the pre-mRNA splicing

Over forty-years after the discovery of pre-mRNA splicing, its fundamental role in RNA maturation as well as its importance in gene expression are indisputable. The progress in structural and global approaches have paved the way for the current understanding of splicing. First described for the adenovirus 2 (Ad2) mRNAs in 1977 (Berget et al., 1977; Chow et al., 1977), splicing is now known to exist as a conserved mechanism of pre-mRNA processing in the majority of eukaryotic organisms (Ast, 2004). This critical step in the maturation of mRNAs involves the recognition and excision of non-coding regions of the gene, called introns, from the pre-mRNA transcripts and subsequent ligation of the flanking coding regions, known as exons (Figure 1.3). The intronic regions of a transcript are determined by three main conserved sequence features

consisting of the 5′ splice site (5′SS), the branch point, and the 3′ splice site (3′SS). In higher eukaryotes the branch point is followed by a pyrimidine rich sequence known as the polypyrimidine tract (Figure 1.3). The 5′SS and the 3′SS define the exon-intron and intron-exon junctions, the polypyrimidine tract assists the 3′SS recognition and the branch point sequence is critical for the first step of the splicing reaction (reviewed in Plaschka et al. (2019)).

The splicing reaction relies on two consecutive phosphoryl transfer reactions. The first reaction is achieved by the nucleophilic attack of the phosphate at the 5′SS by the branch point adenosine 2′-hydroxyl group. This reaction then produces a free 5′ exon and also an intermediate product corresponding to the intron lariat still linked to the downstream 3′ exon. The second reaction involves the attack of the 3′-hydroxyl group from the free 5′ exon at the phosphate in the 3′SS, resulting in the ligated exons and a free intron lariat (Plaschka et al., 2019). Thus the pre-mRNA splicing consists of a simple two-step chemical reaction, yet it is orchestrated by a very complex machinery, the spliceosome.

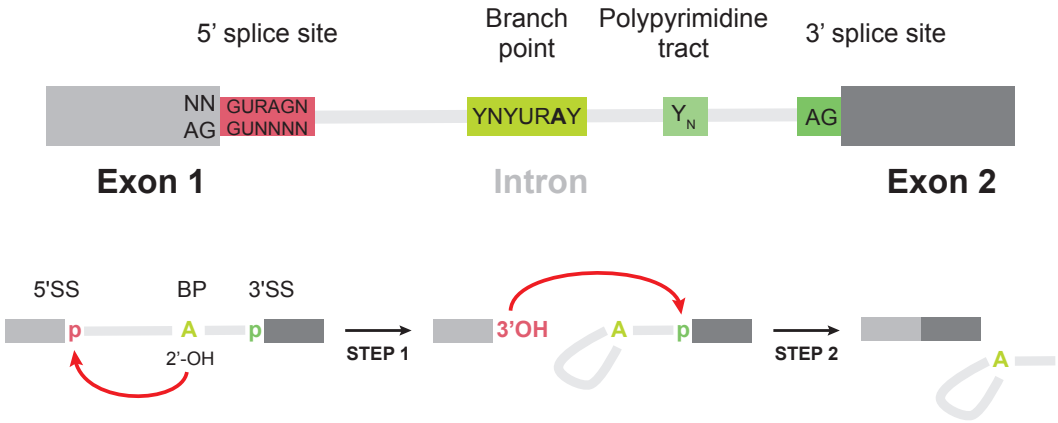


Fig. 1.3 Schematic of the splicing reaction. Top, cis-elements involved in pre-mRNA splicing of humans are highlighted. Conserved sequences at the 5SS and 3SS and the internal branch point (BP) sequence are also shown. **Bottom**, intron removal from pre-mRNAs is catalyzed in a two-step reaction involving phosphoryl transfers. The first step leads to the intron branching and the second one results in the exon ligation. Figure modified from (Plaschka et al., 2019)

1.3.2 The spliceosome

The spliceosome complex is a large ribonucleoprotein which comprises over 100 proteins in humans and is assembled upon the coordinated recognition of the intronic sequences by five major small nuclear ribonucleoproteins (snRNPs): U1, U2, U4, U5 and U6 (Plaschka et al., 2019). These snRNPs, mainly U1 and U2, are important in the initial steps of splicing (pre-spliceosome formation) recognizing and interacting with the RNA elements in the intron defining the sequences to be spliced (Figure 1.4). In fact, the recognition of the splice sites certainly represents a daunting problem, given that the splicing machinery must identify short sequences embedded in very long introns. Therefore, multiple relatively weak interactions and their crosstalk are required to ensure proper recognition of the splice sites before splicing commitment. Formation of this network of interactions across exons and/or introns are termed as exon definition (ED) and intron definition (ID) respectively (Hollander et al., 2016) (Figure 1.4). For instance, binding of the SR proteins (see below) influence binding of U1 snRNP to the downstream 5′ splice site and of U2AF to the upstream polypyrimidine tract establishing a "cross-exon" recognition complex (ED) (Mayeda and Krainer, 1992; Reed, 1996) (Figure 1.4). Nevertheless, intron definition can also be formed by recognition of the pair of splice sites flanking the same intron (Figure 1.4). This mechanism is known to exist in lower eukaryotes, for example in yeast, whose introns are typically shorter than mammalian ones (Hollander et al., 2016).

Recent structural work by cryo-EM (cryogenic electron microscopy) focusing on the yeast spliceosome, due to its simpler splicing machinery, have provided insights into the key steps of the splicing reaction (reviewed in Plaschka et al. (2019)). The splicing cycle involves a series of concerted critical steps that lead to the spliceosome assembly, activation, catalysis of the first reaction with lariat formation, remodeling, catalysis of the second reaction culminating on exon ligation, and final disassembly and recycling (Figure 1.5). Moreover, several states of the spliceosome have been isolated, namely E, A, pre-B, B, B^{act}, B^{*}, C, C^{*}, P, and intron lariat spliceosome (ILS) complexes. The assembly and transitions between the distinct states occur as follows: firstly, a E complex is formed upon the recognition of the 5′ splice site and the branch point by U1 snRNP and the branchpoint bridging protein. Interaction of U2 snRNP at the 3′ splice site forms the A complex, which associates with the pre-assembled tri-snRNP U4/U6.U5 to result in a pre-B complex. The next state of the spliceosome, the pre-catalytic B complex, is achieved upon U1 dissociation via Prp28 helicase activity. Subsequent release of the U4 snRNPs from the complex together with Lsm proteins is carried out by the helicase Brr2. A B^{act} containing the stabilized

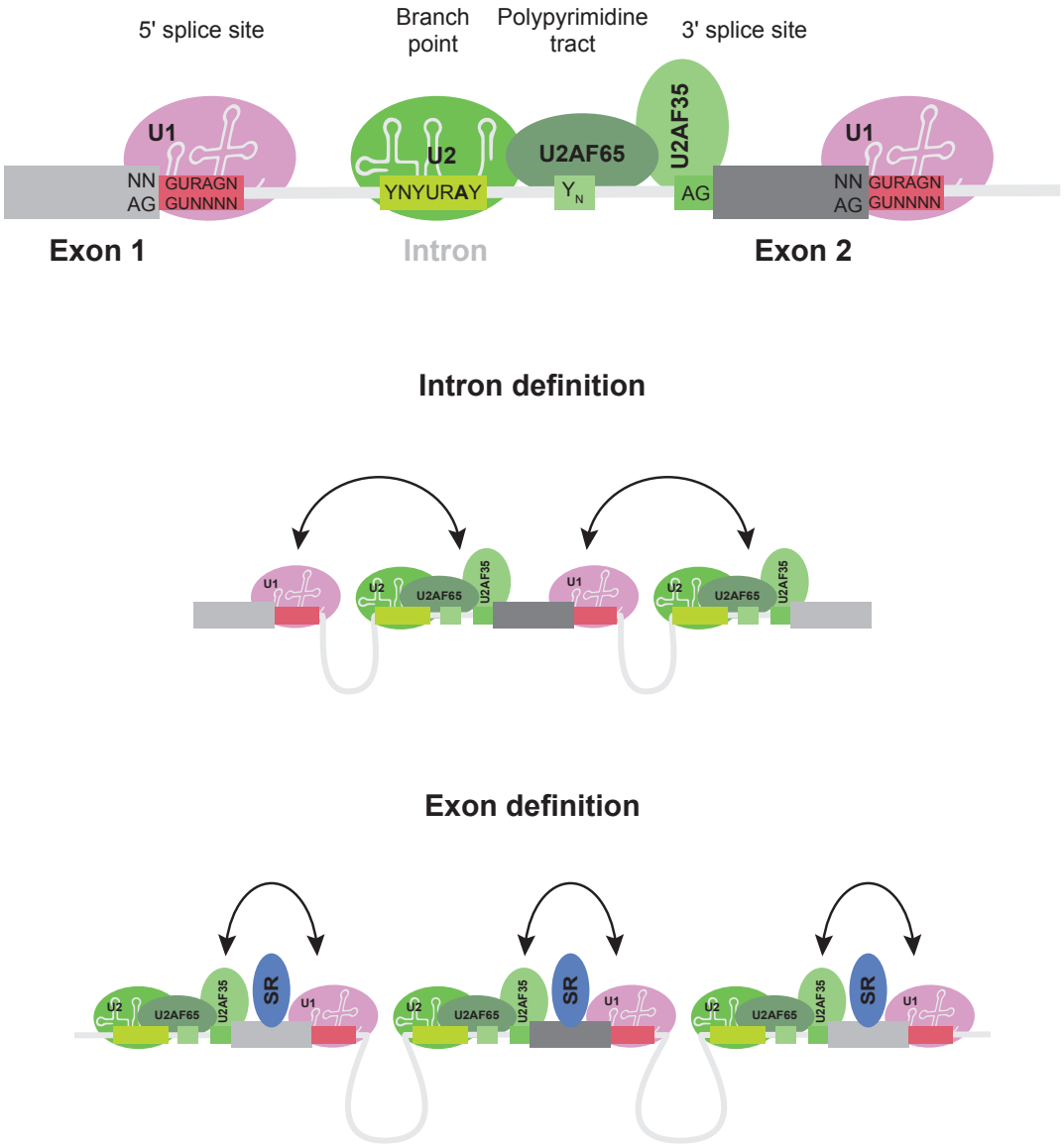


Fig. 1.4 Schematic of the recognition of splicing regulatory sequences. Interactions between RNA elements and snRNPs occurring in the prespliceosome stage. Cross-intron and exon interactions establishing intron definition and exon definition are illustrated at the bottom.

RNA structure of the active site is mediated by the association of the nineteen complex (NTC) and nineteen-related (NTR) proteins. Next, the Prp2 helicase activity drives the dissociation of the SF3a and SF3b from the complex with U2 snRNP and of the retention and splicing (RES) complex. These events lead to the binding of the Prp16 helicase and of the branching factors, Yju2, Isy1, and Cwc25, which bring the branch helix into the active site for the step 1 reaction. Then the nucleophilic attack of the 5′SS by the branch point adenosine 2′-hydroxyl group takes place in the B^* spliceosome stage. As a result of the reaction, a lariat intron intermediate and a 5′ exon are produced with following formation of the C complex. Step 1 factors dissociate together with Prp16 helicase in response to its ATPase activity, further allowing the binding of another helicase, Prp22, as well as the step 2 factors, Prp18 and Slu7. The second reaction is catalyzed in the resulting C^* spliceosome complex. The 3′SS is docked into the active site where the 3′-hydroxyl group of the 5′ exon attacks the 3′SS, generating the lariat intron and the ligated exons. Eventually, the mRNA with the ligated exons is released from the post-splicing P complex by the action of the Prp22 helicase. The remaining ILS complex is disassembled upon Prp43 helicase activity and the recycled snRNPs are then available for a next round of splicing (reviewed in Plaschka et al. (2019)) (Figure 1.5).

Despite the fact that initial studies had been carried out in a non-human system, latest structures from the human spliceosome complex support the remarkable conservation of the spliceosome structure and mechanism between the yeast and human systems ((Bertram et al., 2017; Fica and Nagai, 2017; Yan et al., 2017; Zhang et al., 2017) and reviewed in Plaschka et al. (2019)). Furthermore, even though detailed structural work is very interesting from the point of view of spliceosomal catalysis, it provides little information on splicing regulation. Regulation mostly occurs in the early stages of spliceosome assembly, especially at the level of E complex, which is heterogenous in composition between pre-mRNAs - and also in exon definition complexes, yet all of the structural work is on cross intron complexes (Figure 1.5).

1.4 Alternative splicing

In addition to the discovery of the "split genes", studies of Ad2 mRNAs in the late stage of infection also uncovered the production of multiple mRNAs from a single transcript through alternative choices of exons and splice sites during splicing (Berget et al., 1977; Chow and Broker, 1978; Chow et al., 1977; Nevins and Darnell, 1978). This differential use of exons was then denominated alternative splicing (AS). Soon after, pre-mRNA

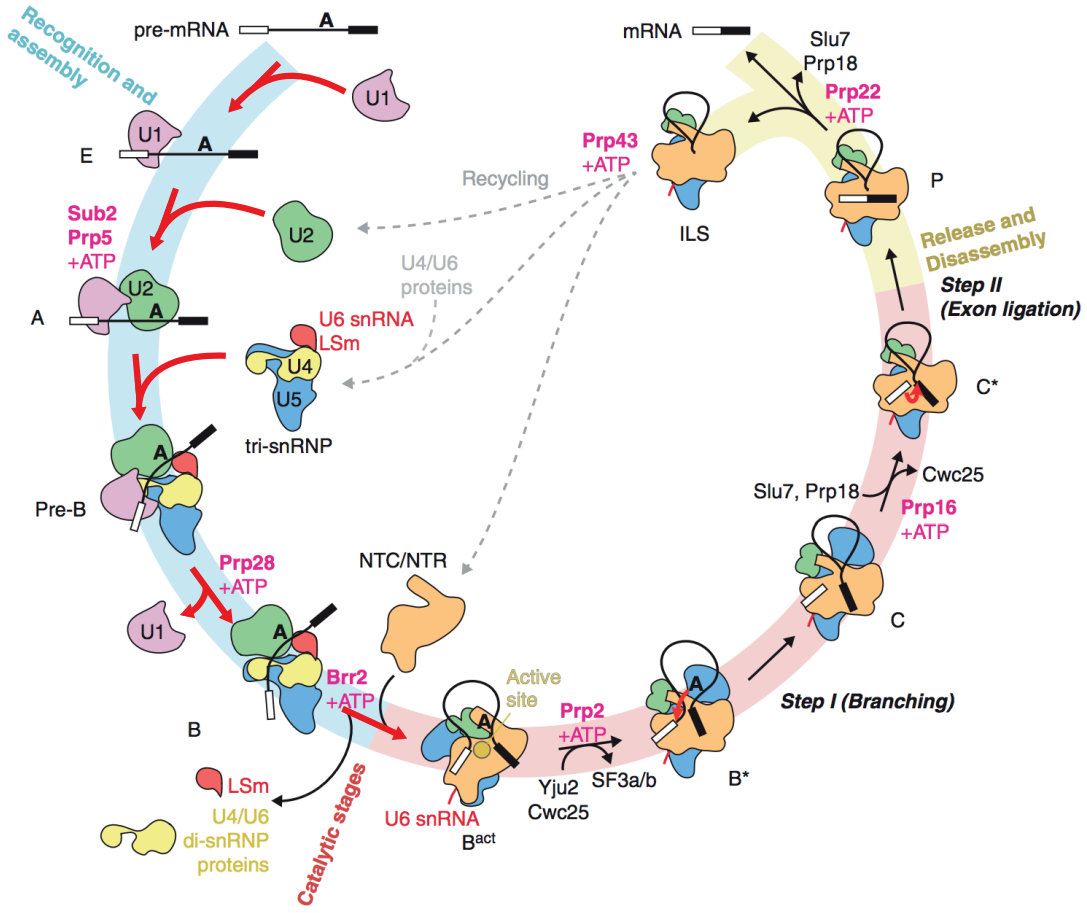


Fig. 1.5 Schematic of the yeast splicing cycle. The pre-mRNA splicing is concerted by a series of steps that can be divided into recognition and assembly, catalytic stages and lastly disassembly and recycling. Several states of the spliceosome have been isolated and are shown in the cycle, namely, E, A, pre-B, B, B^{act}, B^*, C, C^*, P, and intron lariat spliceosome (ILS) complexes. Red arrows indicate stages that are regulated during alternative splicing. Figure modified from (Plaschka et al., 2019).

splicing was shown to be a more general mechanism also acting in the endogenous eukaryotic mRNA processing as revealed by the production of membrane-bound and secreted immunoglobulin by the same gene as well as the two peptide hormones encoded by the single calcitonin gene (Alt et al., 1980; Amara et al., 1982; Early et al., 1980; Rosenfeld et al., 1982). Nowadays, high-throughput sequencing studies have estimated that at least 95% of all the human pre-mRNAs are differentially spliced (Pan et al., 2008; Wang et al., 2008). These studies also highlighted the potential of genes to encode two to thousands of mRNA isoforms. Therefore, it is possible that the generation of multiple isoforms from a single gene by AS is sufficient to explain the paradox of organisms with similar number of genes that yet display completely different levels of complexity (Barbosa-Morais et al., 2012; Sharp, 2005).

AS affects the pattern of multiexon transcripts in many ways, being classified into simple and complex patterns as illustrated in Figure 1.6 (Black, 2003). Simple pattern events can be divided into seven basic types of splicing identified as skipped or cassette exons (SE), mutually exclusive exons (MXE), alternative $5'$ and $3'$ splice sites (A5SS and A3SS), retained intron (RI), alternative promoter and alternative poly(A) sites. Briefly, skipped exons are exons that can either be included or excluded from the mRNA. In mutually exclusive exons, multiple exons are included in an exclusive manner that inclusion of one of the exons imply exclusion of the other one. A5SS and A3SS extend or shorten the length of an exon by controlling the splice site choice. The intron retention pattern results from the not removal of the intron sequence. Finally, alternative promoter selection drives use of an alternative $5'$ end exon and, alternative $3'$ end exons can arise from either alternative splicing or poly-A site selection regulation. Hence, in the transcriptome, the basic patterns are found combined, giving rise to more complex splicing patterns that involve exon choices beyond binary modules (Figure 1.6).

The distinct classes of AS are found in different relative abundance. Interestingly, the AS type landscape also varies among different organisms. In general, exon skipping is the most predominant class whilst intron retention is the rarest in vertebrates, especially mammals. In addition to that, they exhibit more complex events than invertebrates. The accumulation of complex events might be explained by the increase in the number of exons in more complex organisms. Thereby, together with the expansion of the possible splicing combinations, a more sophisticated control of AS has co-evolved in vertebrates (Sammeth et al., 2008).

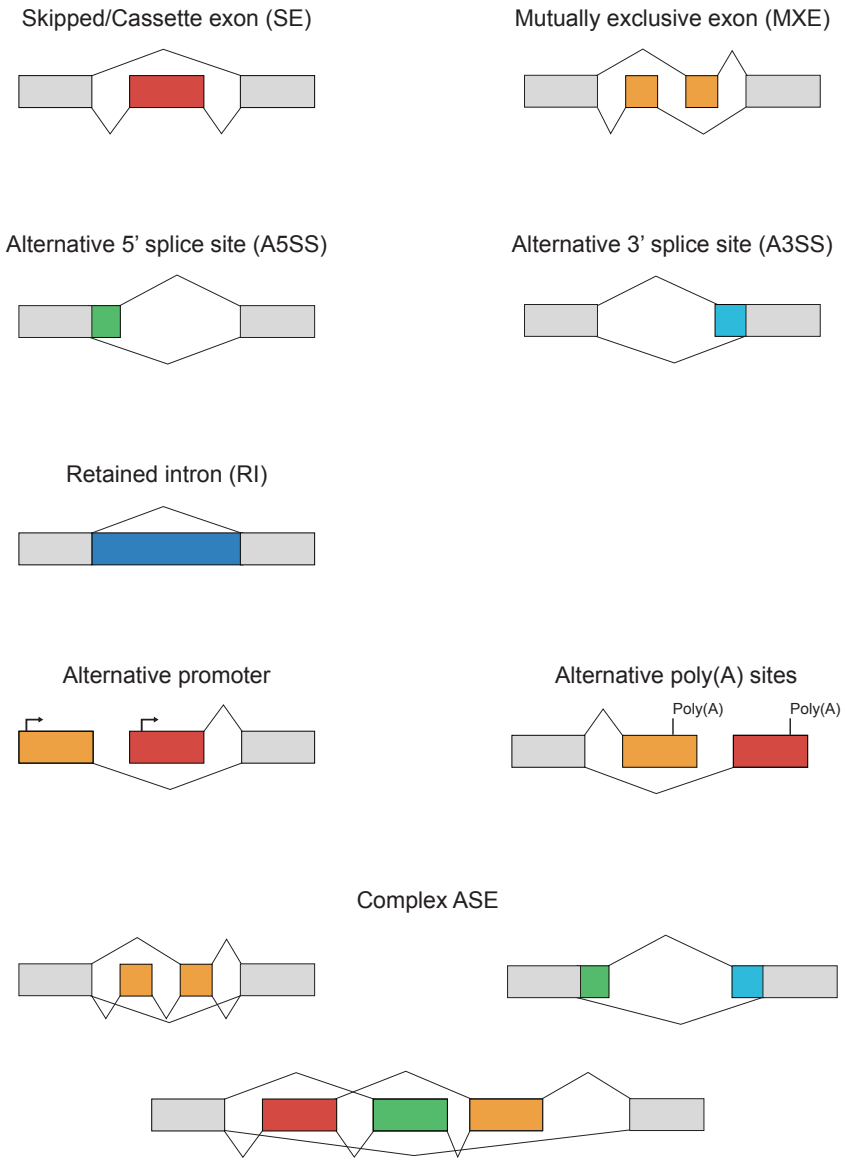


Fig. 1.6 Types of Alternative Splicing. Skipped exon also known as cassette exon (red) can be included or excluded in the transcript. Mutually exclusive exons (orange) consist of two adjacent exons out of which only one of them is included in the final mRNA. Alternative 5′ and 3′ (green and light blue) allow modification of the length of the exon included. Intron retention (dark blue) is defined by the non-removal of an intron from the mRNA. Alternative promoters and poly(A) sites (red and orange) affect the first and last exons of the final transcript respectively leading to alternative transcription start site and polyadenylation. More complex alternative splicing events (ASEs) can arise from the combination of the te different AS types. Constitutive exons are shown in gray and the possible exon junctions indicated by the arches.

1.4.1 Implications of alternative splicing to transcriptome and proteome diversity

The contribution of AS to the molecular diversity is exemplified by the *Drosophila* gene Dscam whose differential splicing could potentially give rise to an extraordinary 38,016 transcript variants (Schmucker et al., 2000). Through alternative choices of exons, AS expands the coding capacity of the genome which is also reflected in the great diversification of the transcriptome and proteome. The inclusion and exclusion of different exons have direct impacts upon the properties of the transcript in the cell, such as localization, stability and translational efficiency. On the other hand, expansion of the proteome is only possible due to the fact that most of the alternatively spliced events tend to maintain the reading frame and therefore be translated into functional proteins (Barbosa-Morais et al., 2012; Merkin et al., 2012; Mockenhaupt and Makeyev, 2015). In that way, the alternatively spliced protein isoforms can have divergent structures and functions consequently affecting their roles in the cell. Interestingly, disordered regions of proteins are more prone to AS regulation as implied by the high enrichment of AS event in these regions. As an effect of that, AS could modulate protein signaling and protein-protein interactions (PPI), known functions of the disordered regions within proteins (Barbosa-Morais et al., 2012). Indeed, AS has been implicated in the rewiring of protein-protein interactions by the regulation of exons encoding binding motifs (Buljan et al., 2012). Furthermore, despite not all splicing events being conserved, genes containing evolutionary conserved alternative exons are likely to encode isoforms with distinct and important functions. For example, the (TGF) beta-activated kinase *Tak1* (MAP3K7) gene has an alternative exon event which has been remarkably conserved in the deuterostome clade (Venables et al., 2012). *Tak1* muscle-specific exon was found used in a tissue-specific manner even in the most ancient member of this group, suggesting a potentially conserved physiological role that requires splicing-specific isoforms (Venables et al., 2012).

Besides generation of protein diversity, AS also allows quantitative regulation of proteins by production of mRNA isoforms that are subjected to Nonsense Mediated Decay (NMD) (McGlincy and Smith, 2008). Indeed, a significant subset of alternative splicing events creates a premature termination codon (PTC) in the mRNA which are then not translated into truncated proteins and instead are directed to mRNA degradation via NMD (Baek and Green, 2005; Weischenfeldt et al., 2012). Therefore, this RNA surveillance pathway guarantees the specific expression of mRNAs that display the correct arrangement of translational signals (Weischenfeldt et al., 2012). Although the generation of mRNAs solely for degradation is arguably a cost of biological regulation,

the biological importance of AS-NMD events is reinforced by their ultraconservation (McGlincy and Smith, 2008).

The functional importance of AS in biological systems is demonstrated by the numerous cases in which regulation of complex physiological processes depends on AS. For instance, generation of different transcript isoforms regulates neurite-growth promoting activity of NCAM, cell aggregation mediated by CD31 and apoptosis regulation by the Fas receptor (reviewed in Kelemen et al. (2013); Stamm et al. (2005)). Moreover, AS-NMD is critical in the control of the RBP PTBP2 during neuronal differentiation (see below) and stabilization of *Myogenin* during myogenesis (Nickless et al., 2017).

Therefore, AS provides biological systems with two post-transcriptional regulatory motifs for establishment of a gene expression network. Firstly by generating multiple protein isoforms and secondly by downregulating gene expression with the coupling of AS and NMD. Eventually, the established network has direct implications on a variety of biological processes in the cell such as proliferation, adhesion, differentiation and death. Thus, AS has the potential to reshape the transcriptome establishing a "robust" network that directly influences cell fate decisions (Jangi and Sharp, 2014).

1.4.2 Regulation of alternative splicing

1.4.2.1 Regulation by regulatory RBPs

Constitutive and alternatively spliced exons are mainly distinguished by the strength of their splice sites (Garg and Green, 2007; Zheng et al., 2005). Constitutively spliced exons normally display strong splice sites that contain the consensus sequences that are recognized by the spliceosome machinery. Consequently, these exons are included in all the transcripts in the cell. Nevertheless, constitutive exons containing non-consensus splice sites are also found and in their case, consistent inclusion occurs due to the presence of secondary splicing regulatory elements. On the other hand, alternatively spliced exons are usually associated with weak splice sites that rely on extra signals for their splicing. In addition to the consensus splicing sequences and core proteins, alternatively spliced exons depend on exonic and intronic splicing enhancer sequences involved in the recruitment of regulatory splicing factors (Garg and Green, 2007; Zheng et al., 2005). The interplay between regulatory *cis* and *trans*-elements determine the final AS outcome of the transcript in the cell.

The *cis*-elements are classified according to their position and function into exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers

(ISEs) and intronic splicing silencers (ISSs). These elements are found encoded in the transcript sequence, making the *trans*-acting factors the main drivers of AS within a specific cell type. The array of *trans*-regulatory proteins involved in the regulation of AS in the cell is comprised of RNA-binding proteins (RBPs). RBPs are proteins that bind to RNA molecules, either single or double stranded, forming ribonucleoprotein complexes that are critical in other post-transcriptional functions beyond AS such as polyadenylation, mRNA stabilization, mRNA localization and translation (Gerstberger et al., 2014). They are evolutionarily conserved and highly abundant in the cell. RBPs exhibit various arrangements of structural domains, sharing common domains but in combination with various auxiliary domains (Figure 1.7). Among the RNA binding domains, the RRM domain is the most abundant domain across RBPs, being present in more than half of the RBPs that bind to mRNA and exhibiting a diverse range of binding motifs (Ascano et al., 2012; Gerstberger et al., 2014).

Fig. 1.7 Structural domain diversity of splicing regulatory proteins. Splicing regulators share common domains yet they are combined in different arrangements. They usually contain RBP domains, such as RRM and KH, combined with other accessory domains. Different domains are color coded. Figure adapted from (Black, 2003).

In spite of the great diversity of domain structures and arrangements of RBPs, RNA splicing maps of different RBPs have revealed general principles underlining

their splicing regulation (Witten and Ule, 2011). These maps are generated by integrating genome-wide protein–RNA interactions studies, such as CLIP (cross-linking and immunoprecipitation) of a single RBP, with the global effects of the same individual RBP on splicing, as profiled by RNA-seq or microarray of overexpression or knockdown (Yee et al., 2019). Therefore, RNA splicing maps are very informative, providing global positioning principles determining splicing regulation by a specific RBP. These RBP-maps highlighted that splicing factors are usually able to repress and activate splicing according to their binding context. In this position-dependent splicing regulation, binding of RBPs to the upstream intron as well as within the exon is associated with silencing effects, whereas binding to the downstream intron enhances splicing. These effects are probably due to the binding near upstream regulatory regions, such as the branch point and splice sites, during repression and via interaction with U1 snRNP or by facilitating the recognition of the 5′ splice site during splicing activation (Witten and Ule, 2011). For example, hnRNP proteins form a large group of RBPs that associates with pre-mRNAs, being able to negatively and positively regulate pre-mRNA splicing according to their binding position (Fu and Ares, 2014). One of its members, hnRNP H, recognizes G-rich sequences and inhibits splicing upon binding to the exon, as shown in the regulation of β-tropomyosin exon 7 (Caputi and Zahler, 2001). Other RBPs such as hnRNP A1, hnRNP F and Nova also inhibit exon inclusion upon binding to exonic sequences (Fu and Ares, 2014). Nonetheless, exceptions to exonic binding-mediated repression have also been reported. For instance, SR proteins bound to elements within the exon promote exon inclusion instead (Mayeda and Krainer, 1992). Moreover, as a result of the positional dependent activity, RBPs can be found promoting opposite effects. hnRNP H when bound to downstream intron is able to activate splicing in contrast with its repressor role mediated by exonic binding. Another example is the RBP PTBP1, whose pyrimidine-rich motifs are enriched downstream of activated exons but within and upstream of repressed exons (Hamid and Makeyev, 2017; Llorian et al., 2010). Additionally, many other RBPs have been found under a similar positional-activity control, e.g. hnRNP l, hnRNP F, NOVA and RBFOX proteins, underlying global principles of splicing regulation by RBPs.

RBP regulation modes that elucidate the mechanism behind the global position-dependent activity are summarized in Figure 1.8. Firstly, repression of splicing upon the binding of RBPs to the upstream intron and to the exon could be explained by the interference with components of the core spliceosome or SR proteins, since the RBP-RNA interaction takes place near critical splicing regulatory sequences, such as the branch point and splice site (Fu and Ares, 2014). RBPs can compete for binding, for

instance in PTBP1 and U2AF65 competition to bind to a regulated polypyrimidine tract (Spellman and Smith, 2006). Moreover, RBPs can also block access of splicing factors by forming multimeric complexes that extend along the RNA in a propagative binding manner. Lastly, RBPs are also able to affect the RNA structure upon interaction with it, hiding the regulatory sequences present on the pre-mRNA (Fu and Ares, 2014; Spellman and Smith, 2006; Witten and Ule, 2011). Secondly, RBPs bound to the intron downstream of the alternative exon might activate splicing by stabilizing the U1 snRNP on the pre-mRNA directly or indirectly via interaction with TIA proteins (Förch et al., 2002; Witten and Ule, 2011). For example, downstream binding of PTBP1 can directly recruit U1 snRNP and activate splicing of *Dtx2* exon 6 (Hamid and Makeyev, 2017). Like repression, changes in the RNA structure by RBPs can expose the $5'$ splice site and facilitate its interaction with U1 snRNP (Witten and Ule, 2011). Additionally, homotypic and heterotypic protein-protein interactions between RBPs could play a role in the exon inclusion by affecting intron definition (Fu and Ares, 2014), as shown for hnRNP A1 or hnRNP H (Fisette et al., 2010). Finally, further studies of other RBPs and RNA splicing maps will shed more light on the global principles of the position-based splicing as well as identify other specific mechanisms that differ from those proposed so far.

Thus, these modes of splicing regulation explain the role of individual proteins in AS decisions, yet the control of an alternative exon consists of a multifactorial system in which the splicing outcome depends on the combinatorial effects from the plethora of regulatory proteins and sequences present in the cell. These different combinations in distinct cell types have been proposed as the splicing code (Figure 1.9). Hence, understanding how precise patterns of regulation affect splicing could allow deciphering this splicing code and eventually predict splicing outcomes in specific cellular and physiological contexts. In fact, the majority of the RBPs are ubiquitously expressed and only 2% of the RBPs are actually found expressed in a tissue-exclusive manner and these could be important in coordinating cell specific functions. Thus, these tissue specific factors are potentially critical regulators in the establishment of tissue-specific AS programs.

1.4.2.2 Other influences affecting alternative splicing

In addition to regulatory RBPs, AS decisions can also be influenced by chromatin state and Pol II elongation rate (reviewed in (Bentley, 2014; Herzel et al., 2017; Luco et al., 2011; Skalska et al., 2017)). These mechanisms of regulation result from the coupling of AS with transcription. Their specific mechanisms and implications are

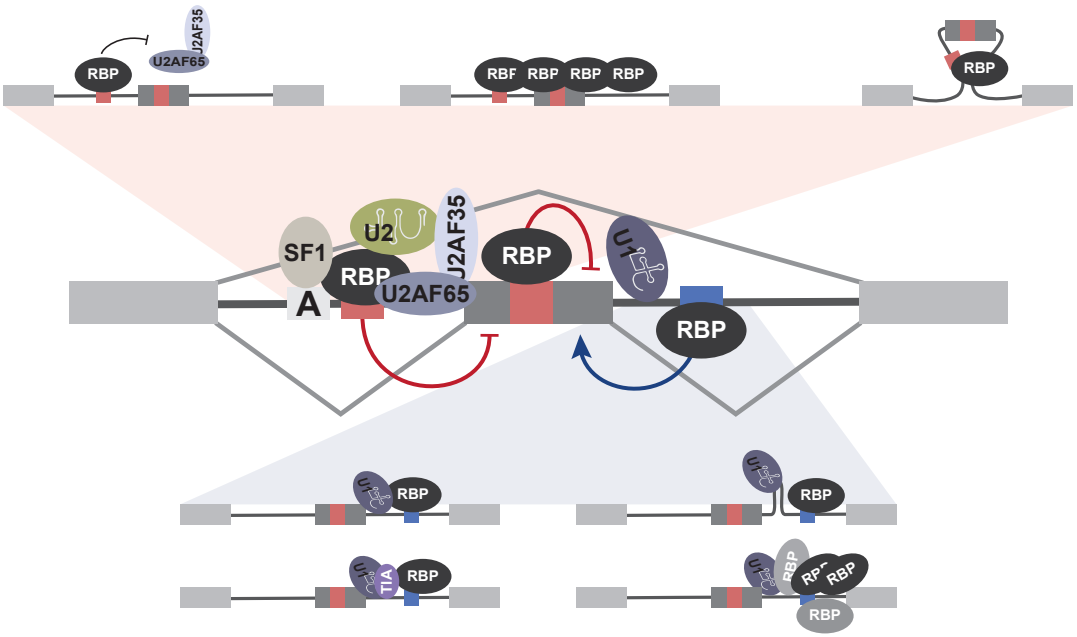


Fig. 1.8 RBP positional binding in the control of AS. Models of the mechanisms underlying RBP positional dependency activity in the regulation of AS. **Top**, models of repression by binding upstream and within the alternative exon. **Bottom**, models of activation by binding downstream of the alternative exon. Constitutive and alternative exons are shown as gray and dark gray boxes, introns as black lines, branch-point sequences indicated by an A and enhancer and silencer sequences as small blue and red boxes respectively.

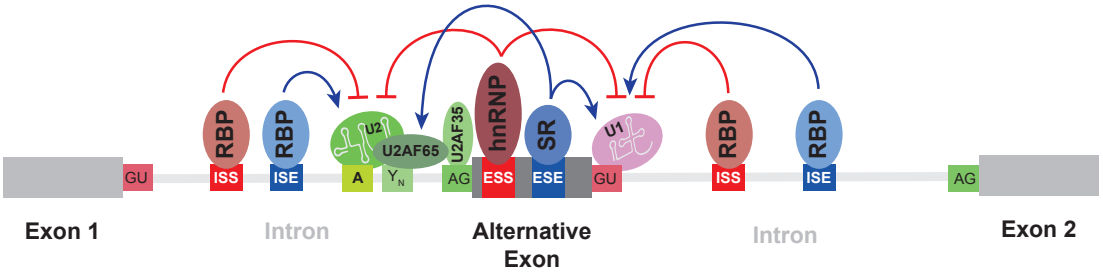


Fig. 1.9 Schematic of alternative splicing regulation. Regulation of AS is driven by protein-RNA interactions between *trans* and *cis* elements. Basic splicing signals present in the pre-mRNA are the 5′ splice site (GU), 3′ splice site (AG), branch point (A), and polypyrimidine tract (Y_n). pre-mRNA *cis* regulatory motifs are represented by exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs), and intronic splicing silencers (ISSs). These regulatory motifs recruit various RNA-binding proteins (RBPs) (e.g., SR and hnRNP proteins). The final splicing outcome is determined by the combination of the individual regulatory signals from the different RBPs. Constitutive and alternative exons are labeled and color-coded as well as the introns. Figure adapted from (Park et al., 2018).

briefly discussed below.

RNA polymerase elongation rates

Splicing mainly occurs during transcription and this opens a window for crosstalk between nascent mRNA, Pol II, and spliceosome, in which Pol II elongation rate, and Pol II CTD modifications are critical factors (reviewed in Herzel et al. (2017); Luco et al. (2011); Skalska et al. (2017)). The contribution of Pol II transcription kinetics to splicing was first hinted in Eperon et al. (1988). In this study, the differences between the *in vitro* and *in vivo* splicing of an A5SS was determined by the length of a loop formed by RNA secondary structure *in vivo*. Thereby, splicing could be indirectly affected by the rate of RNA synthesis through formation of secondary structures in the pre-mRNA. Nevertheless, this assumption was only possible, given the fact that *in vitro* splicing assays utilize pre-mRNAs *in vitro* transcribed that are spliced in a transcription-uncoupled manner. Another piece of evidence for the role of transcription rate in splicing was provided by a study carried in this laboratory. In this study, insertion of a sequence leading to Pol II pausing upstream of *Tpm1* exon 3 regulatory element promoted more inclusion of this exon (Roberts et al., 1998). Moreover, the fact that the transcription rate determined by the nature of the promoter influenced splicing of alternative exons reinforced the coupling of these two processes (Luco et al., 2011).

Therefore, the standing model proposes that slow elongation by Pol II allows inclusion of alternative exons by providing a "window of opportunity" for assembly of the splicing machinery on the alternative exon. On the other hand, Pol II fast elongation rate is more associated with exon skipping due to the competition of the weak splice sites in the alternative exon and the strong splice sites of the downstream constitutive exon already transcribed (Luco et al., 2011). Eventually, Pol II CTD modifications affecting its kinetics can also lead to alterations in AS. Indeed, the splicing of *FN1*, *CASP9* and *BCL2L1* were shown to be regulated upon UV-dependent hyperphosphorylation of the Pol II CTD, which causes a reduction of Pol II elongation rate (Muñoz et al., 2009). Thus, this study highlighted the transcriptional coupling to AS via modification of the Pol II CTD in response to exogenous agents.

The importance of the Pol II rate-regulated AS during mammalian development was recently addressed by generation of mouse embryonic stem cells (ESCs) knocked-in for a slow mutant Pol II (Maslon et al., 2018). Reduction of the elongation rate affected splicing and gene expression in ESCs and resulted in early embryonic lethality in mice and impairment of neuronal differentiation (Maslon et al., 2018). Therefore, it is evident that RNA polymerase elongation rates play a critical role in AS with direct impacts upon the cell fate.

Chromatin context

The studies of epigenetic marks, such as DNA methylation and histone modifications, at the level of transcription have unraveled their role in the organization of accessible chromatin and thereby gene expression (Klemm et al., 2019). Nevertheless, chromatin context has recently emerged as a possible key regulator of AS (Skalska et al., 2017). Initially suggested in view of the coincidence of chromatin modifications with splicing regulatory signals such as splice sites, regulation of AS by epigenetic modifications was then shown to be mediated by recruitment of splicing factors through chromatin remodeler proteins (Luco et al., 2011). Thus, it seems likely that chromatin organizing complexes could facilitate correct assembly of the spliceosome on the nascent pre-mRNA in a Pol II elongation rate independent manner (Luco et al., 2011). Furthermore, a number of splicing events such as *FGFR2*, *FN1* and *NCAM* genes have been found to be sensitive to histone marks (reviewed in Luco et al. (2011)). An interesting mechanism that was then further characterized to directly link these modifications to the splicing machinery was the existence of an adaptor complex (Luco et al., 2010; Sims et al., 2007). For instance, formation of a H3K36me3/MRG15/PTBP1 platform has been described, in which PTBP1 is recruited by the adaptor protein MRG15 which specifically recognizes H3K36me3 (Luco et al., 2010). Therefore, many

more combinations of splicing factors, chromatin-binding proteins, and epigenetic modifications may exist to integrate chromatin context to AS, contributing to the complexity of this post-transcriptional process.

1.4.3 Tissue-specific alternative splicing programs

AS is able to reshape tissue transcriptomes in a qualitative and quantitative manner by generation of specific protein isoforms and regulation of transcript decay via NMD. Indeed, a study of the AS isoforms in human tissue transcriptomes revealed over 60% of alternative splicing events (ASE) to be tissue-regulated (Wang et al., 2008). Moreover, switch-like splicing patterns were also observed, being this a recurrent feature in pair of MXEs (Wang et al., 2008). Thus, this type of AS may be responsible for driving the regulation of exons involved in tissue-specific functions. Surprisingly, SEs that showed a switch-like splicing pattern not only had the exon sequence itself conserved, but also their flanking introns, suggesting the requirement of additional regulatory sequences in their regulation (Wang et al., 2008). Another feature that underlies these specific programs is the expression of tissue-specific RNA-binding factors, for instance RBFOX motifs were enriched downstream of exons more included in tissues where RBFOX proteins are highly expressed (skeletal muscle, heart and brain) (Wang et al., 2008). Consistent with that, RBFOX1 has been shown to be a key factor during neuronal development mediating splicing of a neuro-developmentally important subset of ASEs. Therefore, although tissue-specific regulators represent a marginal fraction of RBPs, they can play a role in the biology of different human tissues via AS regulation (Fogel et al., 2012). Additionally, it has been suggested that a small number of master splicing regulators might control cell-specific programs and these regulators could be identified via the proximity of their genes with transcriptional super-enhancers (Jangi and Sharp, 2014). This approach suggested by Jangi and Sharp (2014) relies on the fact that these large clusters of transcriptional enhancers are associated with genes that control and define cell identity (Hnisz et al., 2013).

Even though individual regulators can be critical to specific AS programs, changes in AS actually respond to multiple factors involving concerted expression of tissue-specific and ubiquitous RBPs (Jangi and Sharp, 2014). During cardiac and skeletal muscle development, for example, a splicing program is achieved by mutual exclusive expression of MBNL and CELF. These RBPs coordinate the splicing transition from embryonic to adult stages by acting antagostically in the regulation of target ASEs (Kalsotra et al., 2008). In agreement with a common feature of tissue-specific splicing programs, other examples of opposing factors in the establishment of splicing networks have been

described such as the ESRP proteins (Epithelial Splicing Regulatory Proteins) and RBFOX during epithelial-mesenchymal transitions (EMT) (Jangi and Sharp, 2014). Therefore, coordination of an ASE by opposing RBPs allows tight splicing regulation and in most cases the switch-like splicing. Additionally, RBPs can also cooperate strongly regulating ASEs, which is the case of RBFOX proteins and PTBP2 in neuronal differentiation (Jangi and Sharp, 2014). Interestingly, these two RBPs are also an example of cross-regulation and signal amplification in splicing cascades. RBFOX proteins increase PTBP2 expression during neuronal differentiation by releasing PTBP2 from its autoregulatory NMD splicing. In that way, RBFOX proteins not only drive AS of its direct targets, they also coordinate secondary splicing changes via PTBP2, another splicing regulator (Jangi and Sharp, 2014).

Therefore, splicing networks are built upon regulation at multiple layers, involving various *cis* and *trans*-factors that act synergistically, antagonistically as well as in cross-regulation. Eventually, all the individual AS effects are integrated in order to regulate a functional coherent set of ASEs characteristic of a specific adult tissue program.

1.5 Smooth Muscle Cells

1.5.1 Smooth muscle cells phenotypic plasticity

Smooth muscle cells (SMCs) are important to the proper function of many physiological systems and organs, such as the gastrointestinal system (stomach and small intestine) ensuring the motility of the digestive tract, as well as the cardiovascular system (blood vessels like aorta) controlling the contraction and dilatation of the vessels (Fisher, 2010; Owens et al., 2004). Thus, as indicated above, the contractility of SMCs is the main property of this cell type, but visceral and vascular SMCs themselves show different contraction abilities. It is prolonged and tonic in vascular SMC (VSMC) whereas in other tissues such as bladder tissue, a more rapid and phasic contraction is observed (Fisher, 2010). An even more distinctive and general feature of SMCs is the fact that they are not terminally differentiated and exhibit phenotypic plasticity unlike other muscles, e.g. skeletal muscle cells (Frismantiene et al., 2018; Owens et al., 2004; Ross, 1993) (Figure 1.10). As a result, SMCs can switch from a differentiated and contractile phenotype to a dedifferentiated and more proliferative phenotype. This interconversion is triggered in response to environmental and signaling cues and leads to modulation of several morphological and physiological features of the SMCs. The differentiated

phenotype is characterized by its typical elongated spindle shape and the high expression of a number of SM-specific contractile proteins, such as ACTA2 and CNN1. Together these characteristics allow the contraction and generation of mechanical output by the SMCs. Conversely, the dedifferentiated phenotype is marked by the expression of genes involved in cell cycle progression and extracellular matrix (ECM) remodeling, like cyclins and collagens respectively. Consequently, the dedifferentiated cells are more proliferative, motile and synthetically active and display a more "hill and valley" morphology instead (Frismantiene et al., 2018; Owens et al., 2004). However, SMCs can revert to the contractile and functional phenotype, which in the case of VSMCs is associated with healthy blood vessels. In this way, the phenotypic modulation consists of a major mechanism by which the SMCs can promote renewing and repairing of the tissue after injury (Frismantiene et al., 2018).

SMC dysfunction has been implicated in a number of pathological situations including cardiovascular diseases, such as atherosclerosis and hypertension (Bennett et al., 2016; Frismantiene et al., 2018), airway diseases like asthma (Panettieri, 2002) and gastrointestinal motility disorders (Sanders et al., 2012). The relevance of the study of SMCs is highlighted by the fact that cardiovascular diseases, only one of the group of disorders affected by mis-regulation of SMC phenotype transition, are the major cause of death worldwide and estimated by the World Health Organization (WHO) to be the cause of death of 17.7 million people worldwide, representing 44% of all deaths in 2015. Taken the biomedical importance of SMCs and the demand for therapeutics, extensive studies have been done. However, the mechanism controlling SMC phenotype remains not fully understood.

When SMCs harvested from human arteries were established in culture, a similar transition from a contractile to a synthetic phenotype was observed (Thyberg, 1996), suggesting that SMC phenotypic switching could be modeled *in vitro*. This paved the way for much of the progress in the understanding of the regulation of this process. Since then, several studies have uncovered the changes in gene expression levels revealing the transcription control network during this process (Frismantiene et al., 2018; Owens et al., 2004; Spin et al., 2012). In addition to that, the transcriptional control mediated by miRNAs during SMC phenotypic modulation has also been intensely assessed (Davis-Dusenbery et al., 2011; Owens et al., 2004; Wang and Atanasov, 2019). These studies have shed some light on some of the mechanisms behind SMC phenotypic plasticity by identifying various regulators of the SMC differentiation such as serum response factor (SRF) and myocardin (MYOCD). SRF is a transcription factor associated with development of other muscle types (skeletal and cardiac) and it has been implicated in

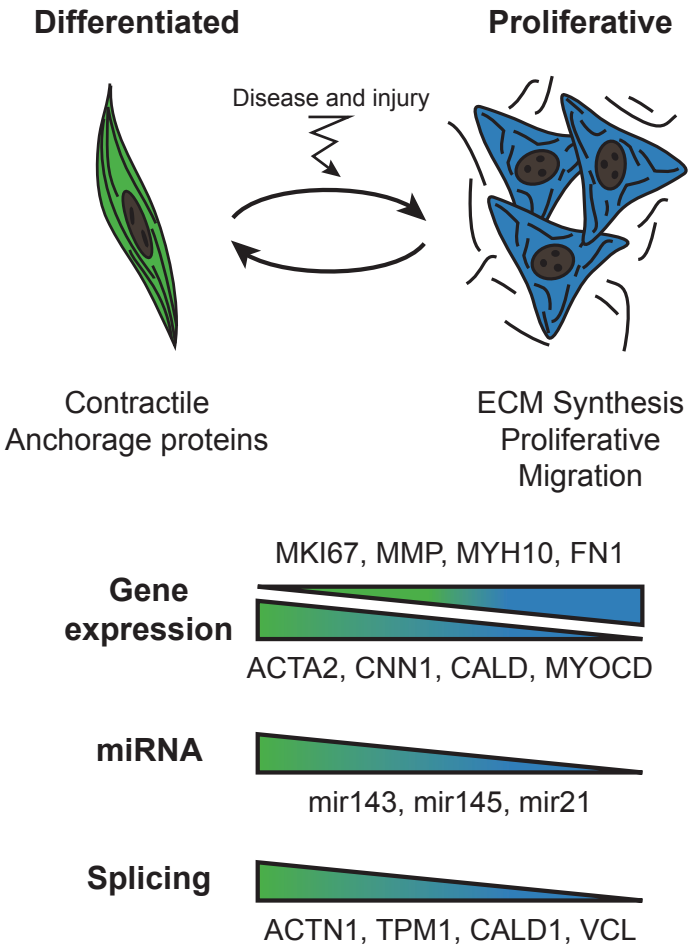


Fig. 1.10 Phenotypic modulation of Smooth Muscle Cells. SMCs present phenotypic plasticity switching between a differentiated (green) and a proliferative (blue) state. Gene expression markers of differentiation and proliferation are indicated as well as miRNAs and genes known to be differentially spliced during SMC dedifferentiation.

SMC specification and differentiation (Sun, 2005). The transcription of many SMC specific genes like SM22α and the α-actin are driven by the CArG boxes (CC[A/T]$_6$GG) located in their promoters which then can be regulated upon SRF binding (Frismantiene et al., 2018; Owens et al., 2004). The latter, MYOCD, is a critical SMC-specific co-activator of SRF and the formation of the complex SRF-MYOCD enhances the transcription activation of the CArG containing genes. In addition to MYOCD, myocardin-related transcription factors (MRTF) can also interact with SRF and act as co-activator of SMC genes. Thereby, disruption of the CArG–SRF–MYOCD/MRTF complex induces the SMC dedifferentiated phenotype. For example, repression of SMC genes can occur by competition between SRF–MYOCD and the co-repressor KLF4 that also binds to G/C elements or displacement of SRF–MYOCD by ELK-1. Moreover, proteins involved in chromatin remodeling also impact activation of SMC genes. However, the outcome of the chromatin remodeling is less clear since both activation and repression roles via modulation of SRF-MYOCD complex and by altering the accessibility to the CArG boxes have been reported. Lastly, miRNAs have been pinpointed as an important regulator of VSMC development and phenotypic switching (Davis-Dusenbery et al., 2011; Wang and Atanasov, 2019). mir-143 and mir-145 were shown to regulate quiescent versus proliferative phenotype in SMCs by targeting critical transcription factors (Cordes et al., 2009). More specifically, these two miRNAs cooperatively regulated mRNA targets, stabilizing *MYOCD* while inhibiting *KLF4* and *ELK-1*, to promote differentiation (Cordes et al., 2009). So, due to the complexity of the SMC differentiation and phenotypic plasticity, it is likely that these biological processes require numerous layers of regulation. However, while the transcriptional program of SMCs has been well-investigated, the knowledge of the regulated splicing programs of SMCs is more rudimentary. The progress in understanding the contribution of post-transcriptional programs in the SMC is discussed below.

1.5.2 The alternative splicing program of differentiated SMCs

Detailed molecular analysis of specific splicing events and global approaches for transcriptome profiling combined with bioinformatic tools have uncovered different AS specific programs across several tissues and stages of development. However, the knowledge of the SMC AS program is still limited compared to other more well-characterized tissue-specific splicing, such as neurons and, skeletal and cardiac muscle (reviewed in Baralle and Giudice (2017); Llorian and Smith (2011)).

Even though some SMC markers of the differentiated state are produced upon differential splicing, e.g. h-Caldesmon and meta-Vinculin (Owens et al., 2004), little is

known about their regulation. In fact, the study of individual SMC-specific ASE has found a number of splicing regulators, such as PTBP1, CELF, MBNL, QKI, TRA2B and SRSF1 (Gooding et al., 2013, 1998; Gromak et al., 2003; Shukla and Fisher, 2008; van der Veer et al., 2013; Xie et al., 2017). Nevertheless, these factor were primarily implicated in the splicing regulation of the proliferative state and were not exclusively expressed in SMCs. Moreover, PTBP1, despite its role in the repression of the mutually exclusive SM exon in the *ACTN1* transcripts (Gromak et al., 2003), also acts in the opposite direction when regulating the splicing of *TPM1* mRNA, facilitating the inclusion of the SM exon by repressing the non-smooth muscle exon, *TPM1* exon 2 and exon 3 respectively (Gooding et al., 1998). Interestingly, the discovery of TRA2B splicing regulation of *Mypt1* brought to attention the differences in splicing within SM tissues (Shukla and Fisher, 2008). *Mypt1* exon 23 was more included in phasic SMCs, in which TRA2B is more expressed. Eventually, this ASE confers unique contractile properties to visceral SMCs allowing their typical fast contraction (Shukla and Fisher, 2008).

Thus, these studies provided interesting functional and mechanistic insights, yet more general regulatory principles of the SM splicing program could not be drawn by them. In an attempt to address that, in the laboratory, the SMC transcriptome was profiled during mouse aorta and bladder de-differentiation using an exon-junction array (Llorian et al., 2016). The global profiling confirmed PTBP1 as a repressor of exons more included in the SM differentiated state (Figure 1.11) and unraveled the concerted regulation of non-productive splicing of post-transcriptional regulators (Llorian et al., 2016). Furthermore, no candidate RBP was identified to regulate the differentiated SMC splicing program as indicated by the motif enrichment (Figure 1.11) and protein expression analysis (Llorian et al., 2016). In fact, it is likely that the identification of critical RBP regulators could be hindered by the rapidly dedifferentiation of SMCs in the cell culture environment. Finally, SMCs are known for their great deal of diversity (Fisher, 2010), even within single blood vessels (Cheung et al., 2012). Therefore, it is possible that distinct RBPs regulate these specialized SMC types as highlighted by the splicing regulator TRA2B.

The findings from the different studies in the SMC regulated AS are summarized in Figure 1.12.

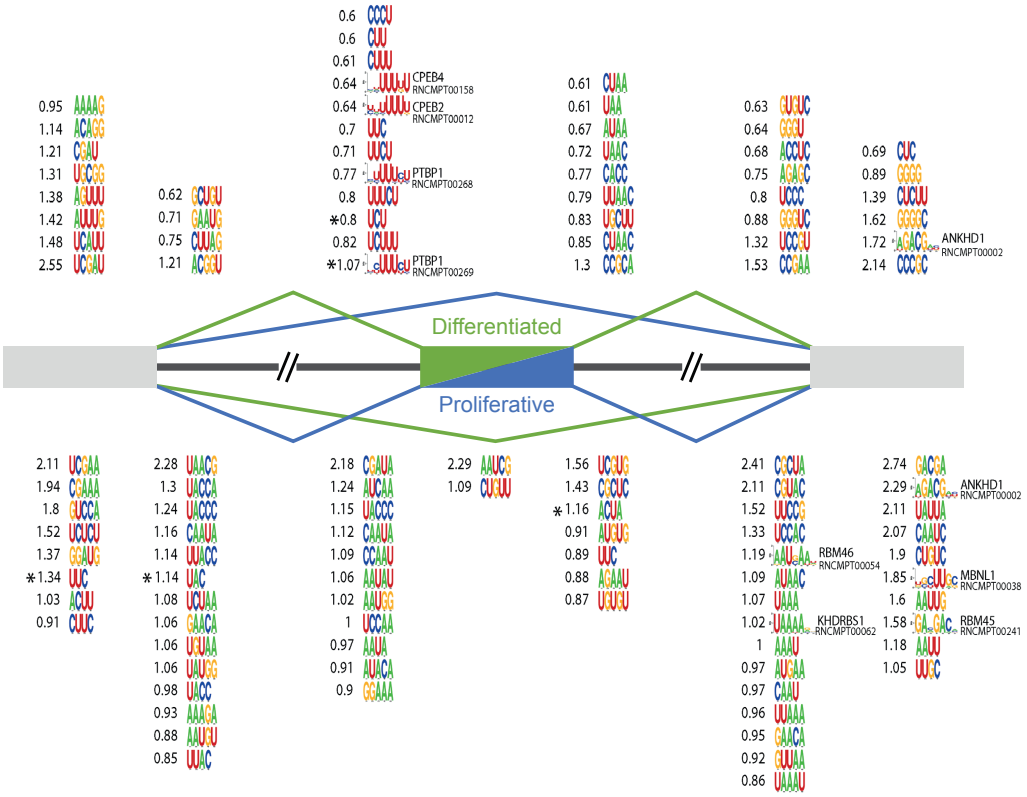


Fig. 1.11 Motif enrichment analysis of SMC regulated ASE. k-mer and RNA complete motifs found enriched in SMC differentiated (green) and proliferative (blue) exons, top and bottom respectively. Motifs shown were significantly enriched ($p < 0.01$). Asterisks indicate motifs with FDR < 0.05 and numbers represent log2 fold enrichment. Figure adapted from (Llorian et al., 2016).

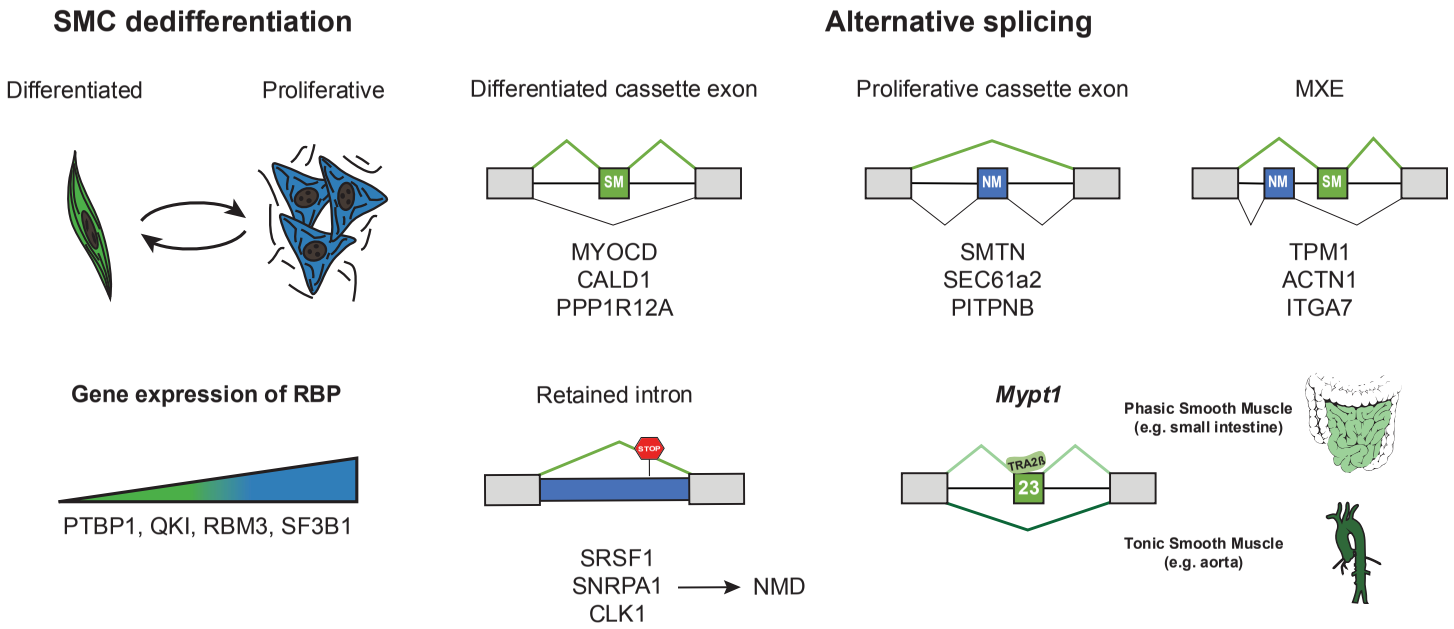


Fig. 1.12 Smooth muscle cell alternative splicing program. Summary of the findings from different studies addressing the SMC regulated AS.

1.6 mRNA splicing in disease and therapeutics

Given the importance of AS in the control of gene expression, it is not surprising that a large number of human diseases are caused by RNA mis-splicing (Table 1.1). Defects in AS regulation due to mutations in *cis*-elements, *trans*-regulatory factors as well as in the core spliceosome components are associated with several diseases such as Duchenne muscular dystrophy (DMD), Spinal muscular atrophy (SMA) and Myelodysplastic syndromes (MDS) respectively (reviewed in Scotti and Swanson (2016)). As introduced before, although splicing occurs in two simple reactions, its proper function requires a concerted series of events operated by the splicing machinery and auxiliary factors. For instance, the 5′ and 3′ splice sites surrounding the right exon must be identified among a large pool of intronic sequences to then drive correct splicing. Therefore, the complexity of splicing makes it susceptible to mutations and genomic sequence variants. Indeed, the Human Genome Mutation Database (HGMD) estimates about 9% of all the reported mutations to affect splicing. Yet this number is underestimated since missense or nonsense substitutions could consist of splicing mutations mainly if lying within regulatory *cis*-elements. In addition to that, about 2000 splice site-disrupting single nucleotide polymorphisms (SNPs) have been identified and are likely to affect splicing (Kurmangaliyev et al., 2013). Moreover, by altering the protein sequence beyond a single residue, splicing mutations have the potential to cause drastic effects. In a way, a single nucleotide mutation within the exon, in spite of being able to cause a change in the amino acid encoded by the codon, can also disrupt regulatory elements that then have the potential to lead to the drastic loss of the complete exon sequence. Additionally, splicing mutations can also result in frameshifting as well as insert premature stop codons that then generate truncated proteins or direct mRNA decay via NMD (reviewed in Montes et al. (2019); Scotti and Swanson (2016)).

A growing number of studies have also indicated mis-splicing events as drivers of cancer progression and contributors of specific cancer hallmarks (El Marabti and Younis, 2018; Montes et al., 2019). Remarkably, cancer cells undergo a large transcriptome reshaping that includes generation of cancer-specific splicing isoforms of oncogenes and tumor suppressor genes (El Marabti and Younis, 2018). It is also notable that all types of cancer cells display striking levels of intron retention, with this type of AS being implicated in the greater diversity of cancer cell transcriptomes compared to normal cells (Dvinge and Bradley, 2015). Additionally, cancer-specific mutations and alterations in splicing factors, changes in signaling pathways controlling splicing factors and, aberrations in core spliceosome components are all linked to deregulation of AS in cancer (El Marabti and Younis, 2018). Finally, a systematic analysis of AS

changes in thousands of tumor samples has underlined a subset of AS events whose consequences have direct implications in the functional transformations characteristic of cancer cells (Climente-González et al., 2017).

The large amount of knowledge gained about the transcript regulatory sequences and auxiliary protein factors as well as their implications to splicing did not only allow the better understanding of the effects of disease-associated mutations but has also contributed to the development of novel therapies based on splicing. The two main strategies being applied consist of antisense oligonucleotide molecules (ASOs) and small molecules (reviewed in Montes et al. (2019); Scotti and Swanson (2016)). The former targets recognition of splicing regulatory sequences in the transcript and its subsequent modulation of splicing whereas the latter has been developed in an attempt to block the binding of *trans*-factors to mutant sequences via regulation of the activity of splicing factors and RNA structures. In fact, an ASO drug has recently been developed consisting of the first medical therapy for spinal muscular atrophy (SMA). The drug, Spinraza™, corrects SMN2 exon 7 splicing by targeting an intronic splicing silencer (ISS) which then induces exon inclusion. SMA is a prominent genetic disease, a cause of infant mortality, and results from the deletion or mutation of the survival motor neuron (*SMN1*) gene, leading to its reduced protein levels (reviewed in Groen et al. (2018)). In humans, the duplicated gene *SMN2* is not sufficient to compensate for the loss of *SMN1* because of its exon 7 predominant skipping (Lorson et al., 1999; Monani et al., 1999). However, defective splicing could be corrected by administration of ASOs, restoring SMN expression in motor neurons after intracerebroventricular injection in mice (Hua et al., 2011). These *in vivo* studies carried out by A. Krainer's laboratory in collaboration with F. Bennett from Isis Pharmaceuticals ultimately resulted in the development of this new splicing-based drug, currently the only treatment approved for SMA (reviewed in Groen et al. (2018)). Therefore, ASOs could be an approach applicable to other mis-splicing disorders. Indeed, pre-clinical proof of concept for another ASO-based targeted therapy have been provided for Familial Dysautonomia (FD), a rare inherited neurodegenerative disorder (Sinha et al., 2018).

Table 1.1 mRNA mis-splicing associated diseases. Table obtained from (Scotti and Swanson, 2016)

Disease	Gene (mutation)	Mechanism	Splicing effect
		Cis	
Limb girdle muscular dystrophy type 1B (LGMD1B)	*LMNA24*	5′ss mutation	Intron 9 retention resulting in NMD
Familial partial lipodystrophy type 2 (FPLD2)	*LMNA25*	5′ss mutation	Intron 8 retention resulting in NMD
Hutchinson-Gilford progeria syndrome (HGPS)	*LMNA26*	Alternative 5′SS	150 nt deletion in exon 11
Dilated cardiomyopathy (DCM)	*LMNA28*	Alternative 3′SS	Extension of exon 4 adding 3 AA to lamin A/C
Familial dysautonomia (FD)	*IKBKAP128*	Decreased U1 recruitment	Exon 20 skipping
Duchenne muscular dystrophy (DMD)	*DMD129*	Exon deletions and skipping	Frameshift resulting in NMD
Becker muscular dystrophy (BMD)	*DMD130*	ESS creation	Exon 31 partial in-frame skipping
Early-onset Parkinson disease (PD)	*PINK1*	U1 5′SS mutation	Cryptic splice site usage, resulting in exon 7 skipping
Frontotemporal dementia with parkinsonism chromosome 17 (FTDP-17)	*MAPT132*	ESS mutation	Increased exon 10 inclusion
X-linked parkinsonism with spasticity (XPDS)	*ATP6AP2*	Novel ESS creation	Increased exon 4 exclusion
		Spliceosome	
Retinitis pigmentosa (adRP)	*PRPF6*	Abnormal nuclear localization	Decreased U4/U6 interaction affecting spliceosome assembly and recycling
	SNRNP200	Decreased helicase activity Decreased proof-reading	Compromised splice site recognition, leading to mis-spliced mRNAs
Myelodysplastic syndromes (MDS)	*U2AF1*	Altered 3′SS preference	Increased alternative 3′SS usage
Microcephalic osteodysplastic primordial dwarfism type 1 (MOPD I)	*RNU4ATAC*	5′ and 3′ stem loop mutations and secondary structure disruption	Compromised minor spliceosome activity
		Trans	
Spinal muscular atrophy (SMA)	*SMN1*	Loss of SMN full-length protein	Altered RNP biogenesis
Amyotrophic lateral sclerosis (ALS)	*TARDP77* *FUS13*	C-terminal mutations alter PPI Decreased U1 interaction Increased SMN binding	TDP-43 target mis-splicing FUS target mis-splicing
Dilated cardiomyopathy (DCM)	*RBM20*	Altered R/S RNA binding domain	*TTN* mis-splicing
Limb-girdle muscular dystrophy 1G (LGMD1G)	*HNRPDL140*	Altered nuclear import of HNRPDL	HNRPDL target mis-splicing
Autosomal dominant leukodystrophy (ADLD)	*LMNB1*	Increased RAVER2 expression	PTBP1 target mis-splicing mediated by RAVER2

During VSMC phenotypic switch in cardiovascular diseases, such as vascular restenosis and atherosclerosis, several genes tightly involved in the phenotype conversion (e.g. *CALD1*, *TPM1* and *SMTN1*) are regulated at the isoform level by differential AS (Llorian et al., 2016; van der Veer et al., 2013). The few studies that have attempted to address the role of AS in VSMCs established a role for the RBP QKI in the dedifferentiated vasculature (van der Veer et al., 2013). The role of RBPs in cardiovascular diseases is not restricted to QKI in VSMC and several other RBPs including MBNL and SRSF1 have been characterized in other cell types of the cardiovascular system (reviewed in de Bruin et al. (2017)). Therefore, better understanding of the *cis* and *trans*-regulatory elements that orchestrate the splicing of critical VSMC genes may, in the future, lead to the development of new splicing-therapeutic approaches in the treatment of cardiovascular diseases.

1.7 Ph.D thesis research aim and outline

To further investigate the regulation of the AS program of differentiated SMCs, the work described in this Ph.D thesis had the following as main questions:

Is there any splicing regulator controlling the SM tissue transcriptome?

If yes, what are the targets it regulates and how does it play its role in alternative splicing?

The answers are divided into five chapters herein (Chapters 3-7). In Chapter 3, the approach suggested in (Jangi and Sharp, 2014) for the identification of potential tissue-specific regulators via association of their genes with transcriptional super-enhancers was applied. This led to the discovery of the RNA-binding protein RBPMS. In Chapter 4, functional studies in SMCs by manipulation of RBPMS levels followed by mRNA-seq were carried out to address any post-transcriptional role of this master regulator candidate. In Chapter 5, bioinformatic analysis combined with *in vitro* and *in vivo* biochemical assays, like *in vitro* binding and minigene transfections techniques, were applied for the confirmation of RBPMS direct regulation of splicing. In Chapter 6, possible indirect effects of RBPMS mediated splicing cascade were further examined, focusing on transcriptional and post-transcriptional processes. Finally, in Chapter 7, further systematic structural-functional studies shed some light onto the mechanisms that could regulate RBPMS activity, e.g. PTM. Eventually, in addition to the summary of the main findings of this work, general discussion and future studies are presented in Chapter 8.

Chapter 2

Materials and Methods

2.1 Identification of potential master splicing regulators

For the identification of super-enhancer associated genes, a catalog of human super-enhancers across different tissue and cell types from (Hnisz et al., 2013) was used. The datasets corresponding to aorta (UCSD_Aorta), bladder (UCSD_Bladder), stomach smooth muscle (BI_Stomach_Smooth_Muscle) and skeletal muscle (BI_Skeletal_Muscle) were compared to identify smooth muscle super-enhancers marked genes. Skeletal muscle was included as an outlier group. To obtain the list of genes associated with each tissue super-enhancer, the dataset tables were loaded into the UCSD Table Browser and data exported by GREAT version 3.0.0. Gene association was determined as follows: "Basal+extension: 5000 bp upstream, 1000 bp downstream, 1000000 bp max extension, curated regulatory domains included". Finally, a list of 1542 human RBPs from Gerstberger et al. (2014) was used for the identification of potential master post-transcriptional regulators.

2.2 Cloning and DNA manipulation

2.2.1 Growth conditions for bacterial cultures

Eschericia coli (*E. coli*) were grown at 37 °C overnight in the presence of the appropriate selective antibiotic. Both LB agar and liquid media were used. For DNA purification, *E. coli* was grown in liquid LB at 37 °C under agitation, using 2 mL and 50 mL LB for small (miniprep) and large scale (maxiprep) purification respectively.

2.2.2 Transformation of *E.coli*

The heat-shock method was carried out for bacterial transformation. Competent cells (TG-1 of DH5a *E. coli*) were pre-made in the laboratory using $CaCl_2$ and stored at -80°C. 10 μL of ligation reaction or 100 ng of plasmid DNA was mixed with 100 μL of competent cells. The mix was then incubated on ice for 30 min followed by heat-shock at 37°C for 2.5 min and another incubation on ice for 2 min. Eventually, cells were spread on LB agar plates containing the appropriate antibiotic. In the case of kanamycin selection, cells were allowed to recover of the heat-shock by addition of 900 μL of LB media and incubation at 37°C for 15 min. Additionally, cells were pelleted and re-suspended accordingly before spreading on LB agar plates.

2.2.3 Ligation of DNA fragments

Ligations were all performed in a final volume of 10 μL and incubated at 4°C for at least an hour. Reactions were set in a 1:5 ratio of vector:insert and using T4 DNA Ligase (Promega). *E. coli* TG1 competent cells were transformed with the ligation reaction as described above.

2.2.4 Preparation of plasmid DNA

For preparation of plasmid DNA by miniprep or maxiprep, single colonies from LB plates were picked and inoculated in LB liquid medium for growth. The bacterial cultures were grown at 37°C for at least 6 hours for minipreps and overnight for maxipreps. Cell were pelleted by centrifugation at room temperature at 8000 g for minipreps and at 4°C at 6000 g for 10 min for maxipreps. Lysis and DNA purification was carried out according to the manufacturer's instructions, QIAGEN QIAprep Spin Miniprep or Maxiprep kit. Concentration of the final purified plasmid DNAs was measured using NanoDrop spectrometer (Thermo Scientific).

2.2.5 Purification of DNA fragments

Fragmented DNAs using restriction enzymes were purified by excision from agarose gel and using QIAGEN QIAquick spin kit. Following steps were performed according to the manufacture's protocol. Final samples were eluted in water and stored at -20°C.

2.2.6 Isolation of PAC1 genomic DNA

To extract genomic DNA from PAC1 smooth muscle cells (Rothman et al., 1992) (Chapter 2.3), first, cells were washed twice with PBS and harvested by scraping off from a single well from a 6-well plate. Cells were resuspended in 50 volumes of 1x proteinase K buffer and 1 volume of proteinase K (10 mg/mL stock) followed by incubation at 37°C for 1 hour. Samples were then phenol extracted and RNase A treated to remove proteins and RNAs from the preparations. Finally, ethanol precipitation was carried to concentrate samples and final elution done with water. PAC1 genomic DNA was then quantified using NanoDrop spectrometer (Thermo Scientific) and stored at -20°C.

2.2.7 DNA constructs

Rattus norvegicus (rat) *Rbpms* isoforms were obtained by PCR amplification from differentiated PAC1 cells cDNA and cloned into pGEM-Teasy (Promega) (see Table 2.2 for oligonucleotide sequences). RBPMS isoforms identified in the PAC1 cells are indicated in Table 2.1. The paralog *Rbpms2* was also cloned (NM_001173426.1/ NP_001166897.1). In order to generate N-terminal Venus and 3xFLAG RBPMS constructs for *in vivo* overexpression, RBPMS isoforms were PCR amplified from the pGEM-Teasy constructs and cloned into the *XhoI* and *EcoRI* sites of the pEGFP-C1 (Clontech) and pCI-neo-3x-FLAG vectors, respectively.

Table 2.1 RBPMS isoforms cloned from PAC1 cells and used in this study

Isoform name	NCBI RNA reference	NCBI protein reference
RBPMS A	XM_006253240.2	XP_006253302.1
RBPMS A.2	XM_006253241.2	XP_006253303.1
RBPMS A.3	not annotated or predicted	not annotated or predicted
RBPMS A.4	not annotated or predicted	not annotated or predicted
RBPMS A.5	not annotated or predicted	not annotated or predicted
RBPMS B	NM_001271244.1	NP_001258173.1
RBPMS B.2	not annotated or predicted	not annotated or predicted

RBPMS mutants were obtained by direct mutagenesis using PCR with oligonucleotides containing the desired mutation, followed by digestion of parental plasmids with *DpnI* enzyme. For generation of C-terminal truncated RBPMS, oligonucleotides containing a stop codon at the position adjacent to the last amino acid codon followed by the specific restriction site were used for PCR amplification. In the case

of N-terminal deletions, oligonucleotides were designed containing the restriction site upstream of the first amino acid codon in the truncated RBPMS.

RBPMS-A and B wild type and RBPMS-A binding mutant were also cloned into a pCI-neo-3x-FLAG containing a nuclear localization signal (NLS) and an MS2 coat protein, both at the C-terminus. Those constructs were then used for *in vivo* overexpression in the tethering assays.

For expression of recombinant RBPMS protein in *E. coli*, the coding sequences of RBPMS-A and B containing an N-terminal 3xFLAG were PCR amplified and then sub-cloned into the *BamHI* and *XhoI* sites of the pET21d plasmid (Novagen).

In the experiments where Venus tagged QKI was overexpressed, the construct used was previously available in the laboratory and it was described in Llorian et al. (2016).

Rat minigene reporter constructs were used to assess splicing activity of RBPMS. The *Tpm1* exon 3 and *Actn1* exon NM-SM were previously available in the lab and their description can be found in Gromak and Smith (2002); Southby et al. (1999). Briefly, the *Tpm1* minigene reporter contains the rat genomic region between *Tpm1* exons 1 and 4, with extensive deletion to intron 1 and 3 and of exon 2. Moreover, the *Tpm1* sequence is flanked by two SV40 sequences in the minigene reporter. The fusion at the 5′ end of exon 1 allows efficient transcription initiation whereas the fusion at the 3′ end of exon 4 drives the processing of the 3′ end of the *Tpm1* minigene reporter encoded transcript. The *Actn1* minigene reporter was constructed in a similar manner, however, the *Actn1* minigene reporter spans over the rat genomic sequence that codes for the pair of mutually exclusive exons, exon NM and exon SM, and both constitutive flanking exons, EF1a and EF2 exons. Moreover, reporters containing only one of the mutually exclusive exons were obtained by deletion of either the SM exon and its flanking intronic regions or the NM exon and some of its flaking intronic sequence as described in (Gromak et al., 2003).

Myocd exon 2a and *Flnb* exon H1 minigene reporters were obtained by PCR amplification of the regulated exons and their flanking introns from PAC1 cells genomic DNA (approximately 250 bp upstream and downstream of *Myocd* exon 2a and 500 bp for *Flnb* reporter; Oligonucleotide sequence in Table 2.3). The PCR products were then transferred into *XhoI*/*EcoRV* and *NotI*/*SphI* sites of the pCAGGs-EGF vector. The pCAGGs-EGFP consists of a GFP expression cassette which has an intron inserted into its second codon. Thus, the *Myocd* and *Flnb* reporters code for an RNA containing three exons and their splicing generate two products that corresponds to the inclusion or exclusion of the exon of interest (Wollerton et al., 2004).

Point mutations of the CAC motifs in the minigene constructs were generated by PCR, using oligonucleotides containing the A to C mutations and the reporter wild-type DNA as template. Moreover, for the tethering assays, MS2 coat binding hairpins were inserted into the CAC mutant *Myocd* and *Tpm1* minigene reporters. Prior to that, the *MfeI* and *NheI* restriction sites were inserted into the CAC clusters locations downstream of the *Myocd* exon 2a and the *MluI* and *BsiWI* restriction sites were included upstream of the *Tpm1* exon 3. Oligonucleotides containing the sequence of the MS2 coat protein binding hairpin flanked by the respective restriction site were used for ligation to the minigene reporters.

For the *in vitro* transcription templates, intronic regions containing CAC motifs from *Tpm1*, *Flnb* and *Myocd* were amplified by PCR and cloned into the *HindIII*/*EcoRI* sites of the pGEM4Z plasmid (Promega).

The sequence of all the constructs were confirmed by Sanger sequencing. Finally, the sequence of all the oligonucleotides used in this study is found in Tables 2.2 and 2.3.

2.3 Cell culture, transfection and inducible lentiviral cell line

The rat pulmonary artery cell line (PAC1) was used as the SMC model in this study (Rothman et al., 1992). This cell line has been shown to be a good system for SMC phenotypic plasticity studies due to its ability to reproduce the differentiated and proliferative states in cell culture. To achieve more differentiated PAC1 cells, they were only passaged once a week. On the other hand, to stimulate the proliferative state, PAC1 cells were passaged twice a week at higher dilution (Llorian et al., 2016).

The human embryonic kidney cell, HEK293, was also used for some of the experiments as an alternative to the poor transfection efficiency of the SMC cell line as well as to address the role of RBPMS in other cell types. Within that same purpose, HeLa cells were used as another human cell line.

All the different cell lines were cultured in DMEM medium containing glutamax and 10% fetal bovine serum (FBS) and under standard conditions. For maintenance, HEK293 and HeLa cells were passaged twice a week in a 1:10 dilution.

For transient overexpression, 1x10^6 PAC1 cells were seeded in 6 well plate a day before the experiment. Cells were then transfected by incubation with a transfection reaction containing 5 μL of the Superfect transfection reagent (Qiagen) and 2 μg of total amount of DNA diluted in OptiMEM. For all the transfections, the amount of

Table 2.2 Sequence of the oligonucleotides used in this study for construction of effectors

Target	5′-3′ sequence	Vector	Cloning method	Experiment
RBPMS A	CGAAGGACCGGGAAGATGAACGGC CTCGAGTCAGCAGAACTGCCGGGA	pGEM Teasy		PAC-1 PCR
RBPMS B	CGAAGGACCGGGAAGATGAACGGC GCTTGCTAATAAATTCAACATGGGAGCTC	pGEM Teasy		PAC-1 PCR
RBPMS2	TCACCATGAGCAACCTGAAGCC GAGGAGCTAACAGAACTGGCGATA	pGEM Teasy		PAC-1 PCR
RBPMS A	CGAAGGACCGGGCTCGAGGTAACGGCGGC CGAGACGAATTCTCAGCAGAACTGCCG	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS B	CGAAGGACCGGGCTCGAGGTAACGGCGGC TTGCTAGAATTCTCAACATGGGAG	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS2	TTCGATTTCCTCGAGGTAGCAACCTG GTGATTGAATTCCTAACAGAACTG	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS A	CGAAGGACCGGGGAATTCAACGGCGGC CGAGACCTCGAGTCAGCAGAACTGCCG	Flag3X	*EcoRI - XhoI*	*in vivo* OE
RBPMS B	CGAAGGACCGGGGAATTCAACGGCGGC TTGCTACTCGAGTCAACATGGGAG	Flag3X	*EcoRI - XhoI*	*in vivo* OE
RBPMS2	GAATTCGATTTCGAATTCAGCAACCTG CTAGTGATTCTCGAGCTAACAGAACTG	Flag3X	*EcoRI - XhoI*	*in vivo* OE
RBPMS A K100E	CGACTAGAGTTTGCTGAGGCAAACACGAAGATG CGACTAGAGTTTGCTGAGGCAAACACGAAGATG	Venus C1	mutagenesis	*in vivo* OE
RBPMS A R38Q	CTGGACATCAAGCCCCAAGAGCTGTACCTGCTCTTC GAAGAGCAGGTACAGCTCTTGGGGCTTGATGTCCAG	Venus C1	mutagenesis	*in vivo* OE
RBPMS A R38A/E39A	CTGGACATCAAGCCCGCAGCGCTGTACCTGCTCTTC GAAGAGCAGGTACAGCGCTGCGGGCTTGATGTCCAG	Venus C1	mutagenesis	*in vivo* OE
RBPMS A	CGAAGGACCGGGGAATTCAACGGCGGC CGAGACACACTCGAGGCAGAACTGCCG	Flag3X+NLS+MS2	*EcoRI - XhoI*	*in vivo* OE
RBPMS B	CGAAGGACCGGGGAATTCAACGGCGGC CTAATACTCGAGACATGGGAGCTC	Flag3X+NLS+MS2	*EcoRI - XhoI*	*in vivo* OE
RBPMS T/E	GAACAAACTCGTAGGGGAGCCAAACCCCAGTGAGCCTCTGCCCAACACTG CAGTGTTGGGCAGAGGCTCACTGGGGTTTGGCTCCCCTACGAGTTTGTTC	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS T/A	GAACAAACTCGTAGGGGCTCCAAACCCCAGTGCTCCTCTGCCCAACACTG CAGTGTTGGGCAGAGGAGCACTGGGGTTTGGAGCCCCTACGAGTTTGTTC	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS A T113E	GAACAAACTCGTAGGGGAGCCAAACCCCAGTACTCCTCTGCCCAACACTG CAGTGTTGGGCAGAGGAGTACTGGGGTTTGGCTCCCCTACGAGTTTGTTC	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS A T118E	GAACAAACTCGTAGGGACTCCAAACCCCAGTGAGCCTCTGCCCAACACTG CAGTGTTGGGCAGAGGCTCACTGGGGTTTGGAGTCCCTACGAGTTTGTTC	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS A TST/E	GAACAAACTCGTAGGGGAGCCAAACCCCGAGGAGCCTCTGCCCAACACTG CAGTGTTGGGCAGAGGCTCCTCGGGGTTTGGCTCCCCTACGAGTTTGTTC	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS A TST/D	GAACAAACTCGTAGGGGATCCAAACCCCGATGATCCTCTGCCCAACACTG CAGTGTTGGGCAGAGGATCATCGGGGTTTGGATCCCCTACGAGTTTGTTC	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS K60R	CATAAAGCTCACATCTCGACAGCCCGTAGGCTTTGTC GACAAAGCCTACGGGCTGTCGAGATGTGAGCTTTATG	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS K109R	ACGAAGATGGCCAAGAACCGACTCGTAGGGACTCCA TGGAGTCCCTACGAGTCGGTTCTTGGCCATCTTCGT	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS A ΔN	AGCGAGGCCACTCGAGGTGAGGAGGAGGTCCGGACC	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS ΔC	CAAGAACAAACTCGTATAGGAATTCAACCCCAGTACTC	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS A Δ20	GGAGGGAGCTCGAGTTACTGGGC	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS A Δ35	GAGTTAGCGCCTGCTCTTTGAGAATTCGCCGCTTTCACC	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS RRM	AGCGAGGCCACTCGAGGTGAGGAGGAGGTCCGGACC CAAGAACAAACTCGTATAGGAATTCAACCCCAGTACTC	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS AB	AAGTCCCGGCAGTTCTGCCAGTGTTTCTCTCCAGAGGCA GCAGAACTGCCGGGACTT	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS BA	GGGAGCTCCCATGTATGCGCTGGCTCCCTC	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS B HSA	CCTGTCTTTTGTCCACCCACTGCAGACCAAC GTTGGTCTGCAGTGGGTGGACAAAAGACAGG	Venus C1	*XhoI-EcoRI*	*in vivo* OE
RBPMS A	AAAAAGCAGGCTCCACCATGGACTACAAAGACCATGACGG AGAAAGCTGGGTTTTAGCAGAACTGCCGGGACTTCCAGCC	pDONR	Gateway cloning	inducible OE line
rec RBPMS A	GCTGCTCGACGGATCCGACTACAAAGACCATG CGAGACACACTCGAGGCAGAACTGCCG	pET21d	*BamHI-XhoI*	rec protein
rec RBPMS B	GCTGCTCGACGGATCCGACTACAAAGACCATG CTAATACTCGAGACATGGGAGCTC	pET21d	*BamHI-XhoI*	rec protein

Table 2.3 Sequence of the oligonucleotides used in this study for construction of minigene reporters

Target	5′-3′ sequence	Vector	Cloning method	Experiment
Myocd minigene	ATGTCTCGAGGAGCAAGCTCAGAATG GGGGGATATCAGAGAGATGTTAGG	pGN-GFP	*XhoI-EcoRV*	*in vivo*
Myocd mCAC1	CTATGGGGAGGCTGCGGGGGGAGGGGACGGGGGG	pGN-GFP	mutagenesis	*in vivo*
Myocd mCAC2	CAGTCGGGGGGGGTAGGGGATGGGGGGGAGAAGC	pGN-GFP	mutagenesis	*in vivo*
Myocd mCAC1 *MfeI*	CTATGGGGAGGCTGCGGCAATTGGGGACGGGGGGATGGCC CTATGGGGAGGCTGCGGCAATTGGGGACGGGGGGATGGCC	pGN-GFP	mutagenesis	*MfeI* CAC1
Myocd mCAC2 *NheI*	GGCTTCTCCCCCCCCATCCGCTAGCCCCCCCCGACTGGCTGCC GGCAGCCAGTCGGGGGGGGCTAGCGGATGGGGGGGAGAAGCC	pGN-GFP	mutagenesis	*NheI* CAC2
Myocd UBE2V1 CAC1	AATTGTACTCACTCACTCTCCACGAGC AATTGCTCGTGGAGAGTGAGTGAGTAC	pGN-GFP	*MfeI*	UBE2V1 mCAC1
Myocd UBE2V1 CAC2	CTAGCTACTCACTCACTCTCCACGAGG CTAGCCTCGTGGAGAGTGAGTGAGTAG	pGN-GFP	*NheI*	UBE2V1 mCAC2
Myocd RNA substrate	GCCAAAGCTTGGACGGGCCATC GTGCCGTGAATTCCTCAGGCAGCC	pG4Z	*HindIII-EcoRI*	*in vitro*
Myocd MS2 hairpin1	AATTGCTAGCGACGATCACGCGCTATC AATTGATAGCGCGTGATCGTCGCTAGC	*Myocd* mCAC	*MfeI*	MS2 tethering
Myocd MS2 hairpin2	CTAGCCTAGCGACGATCACGCGCTATG CTAGCATAGCGCGTGATCGTCGCTAGG	*Myocd* mCAC	*NheI*	MS2 tethering
Flnb	GCGGCCGCCCGTGTTCACATGAACACACATG GCATGCAATGTCTGAGTCCCAAGAGGC	pGEM-Teasy		*in vivo*
Flnb minigene	GCAGCTGTAGCGGCCGCCCGTGTTCACATG GACTAGGTCGCATGCAATGTCTGAGTCC	pGN-GFP	*NotI-SphI*	PAC1 PCR
Flnb mCAC1	GTCCCCCCCCCCGATCCCCTCCTCATCCCAACCCCCCTCTCTCTCTC GAGGCTCCACACCTAGAAAAGG		mutagenesis	*in vivo*
Flnb mCAC2	GTAAGATCCCCCTCATGTCCCCCCCTGCTCCCCCCGAATCGCCC GAGGCTCCACACCTAGAAAAGG		mutagenesis	*in vivo*
Flnb mCAC3	GCCTGCTACCCCTCTGCCCCTTTAAAACCCCCCTTCTTTTCCAGGC GAGGCTCCACACCTAGAAAAGG		mutagenesis	*in vivo*
Flnb RNA substrate	CCTGGCCAAGCTTGTGTCCC GTTGGAATTCGCCTGGAAAAG	pG4Z	*HindIII-EcoRI*	*in vitro*
Tpm1 exon 3 mCAC	GGCGCGCGGTGTGGCCCTGCCCACGAATGGCTAACTTTC GAAAGTTAGCCATTCGTGGGCAGGGCCACACCGCGCGCC	pT2	mutagenesis	*in vivo*
Tpm1 exon 3 mCAC	CCTTTATGGTCTACGCCCCCTCAACCCGCCCCTTGCGGGATCACG CGTGATCCCGCAAGGGGCGGGTTGAGGGGGCGTAGACCATAAAGG	pT2	mutagenesis	*in vivo*
Tpm1 exon 3 mCAC	GCCCCTTGCGGGATCCCGCTGCCTGCTGCATC GATGCAGCAGGCAGCGGGATCCCGCAAGGGGC	pT2	mutagenesis	*in vivo*
Tpm1 exon 3 mCAC	CTGCCTGCTGCATCCCCCCCCCTTCCCCCTTC GAAGGGGGAAGGGGGGGGGATGCAGCAGGCAG	pT2	mutagenesis	*in vivo*
Tpm1 exon 3 mCAC	CCCCCTTCCTTCCCCCCCCCCGTACTCCCCTGCCAACTCCCAGC GCTGGGAGTTGGCAGGGGAGTACGGGGGGGGGGAAGGAAGGGGG	pT2	mutagenesis	*in vivo*
Tpm1 RNA substrate	TCCACTAAGCTTCCGGGCGCGCG CAGAGATGCTACGTCAGCTTCAGC	pG4Z		*in vitro*
Tpm1 MluI CCC1	CTGTCCTTTATGGTCTACGCGTCCTCAACCCGCCCCTTG CAAGGGGCGGGTTGAGGACGCGTAGACCATAAAGGACAG		mutagenesis	MS2 hairpin
Tpm1 BsiWI CCC2	CCCTTCCTTCCCCCCCCCCGTACGCCCCTGCCAACTCCC GGGAGTTGGCAGGGGCGTACGGGGGGGGGGAAGGAAGGG		mutagenesis	MS2 hairpin
Tpm1 MS2 hairpin CCC1	CGCGTCTGTGCACGATTACGGCACATGCGA CGCGTCGCATGTGCCGTAATCGTGCACAGA	pT2 mCAC	MS2 hairpin	MS2 tethering
Tpm1 MS2 hairpin CCC2	GTACGCTGTGCACGATTACGGCACATGCGC GTACGCGCATGTGCCGTAATCGTGCACAGC	pT2 mCAC	MS2 hairpin	MS2 tethering

reporter DNA used was 200 ng whereas for RBPMS constructs, the DNA was titrated to achieve similar levels of expression across isoforms. To keep the amount of DNA constant, pGEM4Z (Promega) DNA was added to the transfection reactions.

Transfections for transient overexpression in the human cell lines, HEK293 and HeLa, were achieved in a similar way to the PAC1 cells, although here the transfection reagent Lipofectamine2000 was used instead. $2x10^5$ cells were plated and incubated with 2 μL of Lipofectamine2000 and 2 μg of DNA diluted in OptiMEM.

Rbpms depletion in differentiated PAC1 cells was carried out by siRNA knockdown. $1x10^5$ differentiated PAC1 cells were plated on a 6 well plate and then transfected with Oligofectamine reagent (Invitrogen) and 90 pmol of either the control siRNA (C2: AAGGUCCGGCUCCCCCAAAUG, Dharmacon) or the *Rbpms* siRNAs (siRNA1: RSS363828, GGCGGCAAAGCCGAGAAGGAGAACA , siRNA2: RSS363828, CGCUUCGAUCCUGAAAUCCCGCAAA, Stealth, Thermo Fisher Scientifics). After 24 hours, cells were treated one more time with siRNAs, using Lipofectamine2000 and siRNAs at the same concentration used in the first treatment. To assess any potential cross regulation between *Rbpms* paralogs, an *Rbpms2* knockdown was performed alongside a double knockdown, *Rbpms* and *Rbpms2*, using the following *Rbpms2* siRNA: RSS367260, AAGCCUAGAGAACUCUACCUGCUUU (Stealth, Thermo Fisher Scientifics).

In all the transient overexpression and knockdown experiments, RNA and protein samples were harvested 48 h after transfection.

During the period in which I was establishing the *Rbpms* knockdown experiments, Clare Gooding in the laboratory generated inducible lentiviral *Rbpms* overexpression cell lines. Firstly, C. Gooding cloned 3xFLAG rat RBPMS A cDNA into pINDUCER22 (Meerbrey et al., 2011) using the Gateway system. Lentiviral particles were then produced by transiently transfecting $2x10^6$ HEK293T cells with a combination of 0.75 μg of each of the lentiviral packaging plasmids (gag, pol, tat and VSV-G), 7 μg of the pInduce22 plasmid and 30 μL of Mirus TransIT-lenti (MIR6604) transfection reagent. The medium was collected 24 h after transfection and replaced with fresh medium. After a further 24 h the medium was again collected and pooled together with the first one, which had previously been stored at 4°C. Pooled medium was then centrifuged at 1000 g for 5 min and filtered through a 0.45 micron PVDF filter. For transduction of PAC1 cells, $1x10^4$ proliferative PAC1 cells were plated on a 6 well plate and 24 h later, medium was replaced with the lentiviral medium diluted 1:2 with fresh medium containing 16 μg/mL polybrene. Cells were incubated with lentiviral particles for 24 h and medium replaced by fresh medium. Transduced cells were further expanded as necessary. Lentiviral control populations were obtained by incubating

proliferative PAC1 cells with medium containing lentiviral particles lacking RBPMS-A cDNA. Finally, overexpression was induced with 1 μg/mL doxyclycline in medium for 24 h when RNA and protein were harvested.

2.4 Western blotting

Protein levels in depletion and overexpression experiments were verified by western blotting. Total lysates were harvested by direct addition of protein loading buffer to the cells followed by nucleic acid digestion with 1 μL nuclease benzonase (Millipore). Lysates were then run on a SDS-PAGE gel and transferred to Immobilon-FL membrane (Millipore). The primary antibodies used for detection of specific proteins and loading controls in this study are listed in Table 2.4. Moreover, donkey anti-mouse and anti-rabbit IgG (Stratech Scientifica) antibodies conjugated with peroxidase were used at a 1:10.000 dilution. Luminol/Iodophenol chemiluminescence system was carried for detection using Fuji Medical X-ray film. Films were then developed and digitalized.

Table 2.4 List of antibodies used in this study

Gene name	Antibody	Concentration
ACTA2	Agilent/Dako, M0851	1:500
LSM14B	Sigma Aldrich, HPA041274	1:1000
MBNL1	In-house polyclonal antibody	1:1000
RBPMS	Sigma Aldrich, HPA056999	1:500
FLAG	Sigma Aldrich, F-1804	1:2000
GFP	Invitrogen, A-11122	1:1500
GAPDH	Santa Cruz, sc-25778	1:1000
TUBULIN	Abcam, ab6160	1:5000

2.5 Immunostaining

Immunodetection of RBPMS in differentiated and proliferative PAC1 cells was carried out by growing 5x10^4 cell on coverslips. After 24 h, cover slips were incubated in 4% paraformaldehyde (PFA) for 5 min and rinsed with phosphate buffer saline (PBS). To permeabilize cells, cover slips were transferred to 0.5% NP-40 for 2 min and then rinsed with PBS. Subsequently, cells were incubated for 1 h with 1% bovine serum albumin (BSA) in PBS blocking buffer which was then replaced with RBPMS antibodies diluted in blocking buffer (1:50) and incubation carried for another hour. Primary antibodies

were removed by PBS washes and coverslips incubated for 1 h with anti-rabbit Alexa Fluor 647 secondary antibodies (1:500 in blocking solution). After washing off the secondary antibodies with PBS, coverslips were mounted on ProLong Diamond Antifade with DAPI (Thermo Fisher Scientific). All the washes and incubations were conducted at room temperature. For visualization and image acquisition, a fluorescence microscope (Zeiss Ax10, 40X) attached to CCD AxioCam was used. Finally, acquired images were analyzed on AxioVision V4.8.2 software.

2.6 RT-PCR and qRT-PCR

Total RNA was extracted using TRI reagent (Sigma) according to manufacturer′s instructions. Then, RNAs were DNase (Ambion) treated and quantified using a NanoDrop spectrometer (Thermo Scientific). Next, cDNAs were synthesized and eventually used to monitor changes in gene expression and splicing by PCR and qRT-PCR as described below.

cDNA was prepared with 1 μg of total RNA, oligo dT or gene specific oligonucleotides (Table 2.6) and SuperScriptII (Thermo Fisher Scientific) according to manufacturer′s protocol. Minus RT reactions lacking the reverse transcription enzyme were also made to be used as negative controls.

In order to assess gene expression levels, qRT-PCRs were carried out using 50 ng of cDNA, specific oligonucleotides for target gene (Table 2.5) and SYBER Green JumpStart Taq Ready Mix (Sigma). For the quantitative PCR, the three-step amplification protocol was carried out in a Rotor-Gene Q instrument (QIAGEN). Finally, comparative quantitative analysis was performed in the Rotor-Gene Q Series Software 1.7. Relative expression values were obtained by normalizing the quantification of the target gene to the geometric mean of the quantification of two housekeeper genes, *Gapdh* and *Rpl32*.

Table 2.5 Sequence of the oligonucleotides used in this study for qRT-PCR

Target	Species	Forward (5′-3′)	Reverse (5′-3′)
Rbpms	*R. norvegicus*	AAAGCCGAGAAGGAGAACACC	TTGAATGGCCTGAAGAGCAG
Rbpms2	*R. norvegicus*	TACGAAGGGTCCTTGATCAAGC	TGCGTTTTTGGCTGCTTCTG
Acta2	*R. norvegicus*	TGAAGAGGAAGACAGCACAGC	AAACAGCCCTGGGAGCATC
Cnn1	*R. norvegicus*	ACAGATCAACCCCTGGATCAG	TTGAGCGTGTCACAGTGTTC
Smtn	*R. norvegicus*	CACGCAAGGCCATGATTGAG	ATCTGCTTGATGCTGTTGGC
Gapdh	*R. norvegicus*	GCCTTCTCTTGTGACAAAGTGGA	CCGTGGGTAGAGTCATACTGGAA
Rpl32	*R. norvegicus*	GCCCAAGATCGTCAAAAAGAGG	ATCAGGATCTGGCCCTTGAATC

Table 2.6 Sequence of the oligonucleotides used in this study for minigene reporter PCR

Minigene Reporter	RT oligonucleotide (5′-3′)	Forward (5′-3′)	Reverse (5′-3′)
Tpm1 and *Actn1*	GCAAACTCAGCCACAGGT	GGAGGCCTAGGCTTTTGCAAAAAG	ACAAAAACTCACTGCGTTCCAGGCAATGCT
Myocd and *Flnb*	TAGTTGTACTCC	TTGGCAAAGAATTCGCCACCATGGT	CGTCGCCGTCCAGCTCGACCAG

To detect splicing isoforms, PCRs were prepared with 50 ng of cDNAs, specific oligonucleotides (Table 2.6 and 2.7) and Jumpstart Taq DNA polymerase (Sigma). For visualization and quantification of the splicing isoforms, PCR products were separated in a QIAxcel Advanced System (QIAGEN) and the percentage of spliced in (PSI) value calculated with QIAxcel ScreenGel software. In addition to the minus RT reaction, a no template reaction was also included in all the experiments. PSIs are herein shown as the mean (%) ± standard deviation (sd) of three experiments, unless otherwise specified in the text and/or figure legend.

Statistical significance was tested by two tailed unpaired Student t-test for all the experiments except for the RBPMS-A inducible overexpression where a paired Student t-test was used. p-value is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

2.7 RNAseq and analysis

To globally evaluate RBPMS regulation of mRNA abundance and splicing, an RNAseq experiment was performed for RBPMS depletion and overexpression in PAC1 cells. Total RNA was harvested from three biological replicates of RBPMS knockdown in differentiated PAC1 and three populations of proliferative PAC1 cells overexpressing RBPMS-A and their respective controls. Thereby, the comparison between the knockdown differentiated PAC1 control and the overexpression proliferative PAC1 control provided insights into the mRNA changes in the biological context of the SMC phenotypic plasticity in the PAC1 cell line.

PAC1 cells were lysed with TRI reagent and total RNA extraction carried out by Direct-zol purification column (Zymo research). RNAs were DNase treated (Ambion) to remove genomic DNA traces from the preparations and next quantified using NanoDrop spectrometer (Thermo Scientific). polyA RNAseq libraries were then prepared using NEBNext RNA library kit by the NGS facility in the Wellcome Trust - Medical Research Council Stem Cell Institute where the sequencing was also conducted on an Illumina HiSeq4000 platform. Libraries were multiplexed across two lanes and sequenced on a 150 bp paired end mode to generate approximately 60 million reads per sample. Both RNA samples and RNAseq libraries had their qualities assessed prior to their use.

Table 2.7 Sequence of the oligonucleotides used in this study for splicing PCR

Gene name	Genome	Regulated Exon	5′-3′ sequence
Flnb	hg19	Chr3:58127584-58127623(+)	ATCGCCTCCACTGTGAAAAC
			AATCCCAGGCCGTTCATGTC
Mprip	hg19	Chr17:17083920-17083983(+)	ACCCGGCAACTCAGAAACATC
			AGCTTCAACCGTTCTTGCAC
Tpm1	hg19	Chr15:63335905-63336030(+)	GAGCTGGCAGAGAAAAAG
			AGGAGGACATCGCGGCCA
			TCCAACTCTTCCTCAACCAG
Actn1	hg19	Chr14:69345705-69345786(-)	ATCAGCCAGGAACAGATG
			ACATGAAGTCAATGAAGGCYTG
Ncor2	hg19	Chr12:124811955-124812179(-)	ACACCCACAACCGGAATGAGCCTG
			GGACTTGGCTTTTCGGCTGCTG
Itga7	rn6	Chr7:3363146-3363278(+)	GGGGAGTGGAAGTTCTGTGA
			CCTTCCCAGAATCAATGGAG
Tpm1	rn6	Chr8:72840038-72840164(-)	AGGAGGACATCTCAGCAA
			GAGCTGGCGGAGAAAAAG
			CCAACTCCTCCTCAACCAG
Actn1	rn6	Chr6:103379819-103379900(-)	ATCAGCCAGGAACAGATG
			ACATGAAGTCAATGAAGGCYTG
Flnb	rn6	Chr15:18780269-18780341(-)	CAGGGAAGGGGAAAGTCACC
			ACTCACTGGGACATAGGCCT
Piezo1	rn6	Chr19:55313726-55313798(-)	CAACTCCAGTCCACAGACCC
			TTCCTCCTCACTGTCCGACT
Ptprf	rn6	Chr5:137046364-137046397(-)	CGTCAAAGGATGAGCAGTCAATC
			CCCATCATTGGCTTTGAGAC
Hspg2	rn6	Chr5:155853492-155853543(+)	CTGGGGGTTCAAGTTCCGAC
			CGTGCAGACTCTTGGGAACT
Ppfia1	rn6	Chr1:217680291-217680366(-)	GCCAGTTGCAAGAACGTCTG
			TTGCCATGTCTCTCCTCAGC
Ppfib1	rn6	Chr4:181394960-181394993(+)	GACGAAAGGAGAAGGGGTCG
			CCATCAGAGACTCCACTGCC
Mbnl1	rn6	Chr2:150864085-150864121(+)	AGCTGTACTTCCCCCATTGC
			TGGCTAGTCAGATGTTCGGC
Mbnl2	rn6	Chr15:105772324-105772403(+)	GCCTTCCCTCCCGGTGCTCTTC
			GCAGATTCTTGGCATTCCATTCC
Myocd	rn6	Chr10:51733970-51734014(-)	CAGTTACGGCTTCAACAGAGAA
			TTTTCCTCGGGTCATGGAAC
Arhgef7	rn6	Chr16:83023988-83024213(-)	GAAGGTCACGTCTGTGAGCAACC
			GTCTTGGGACTCCTGCTGAG
Smtn	rn6	Chr14:83758453-83758596(-)	TTCTTCCCTGAGGCTTTTGA
			TGGTTGTACAGCGATTGCAC
Cald1	rn6	Chr4:62260091-62260175(+)	CAGGATGCTGAAGACAAAAAGA
			AGGCACAAGACGACAGACTG
			TGTGGGTCATGAATTCTCCA
Lsm14b	rn6	Chr3:175415535-175415613(+)	GCTAAGAAGCTGTTGCCCAG
			GCGATTTTGCCCTCTGGAAC
Ncor2	rn6	Chr12:37030871-37031101(+)	ACACCCATAACCGGAATGAGCCTG
			GGACTTGGCTTTTCGGCTGCTG

Initial RNAseq analysis as well as the mRNA abundance and splicing analysis described below were performed by Dr. Miriam Llorian. Briefly, Trimmomatic version 0.36 (Bolger et al., 2014) was used for read trimming and adapter removal. Then reads were aligned to the Ensembl rat genome Rnor_6.0 release 89 by using STAR version 2.5.2a (Dobin et al., 2013). Finally, RSEM package version 1.2.31 (Li and Dewey, 2011) was used to perform gene level counts.

2.7.1 mRNA abundance analysis

DESeq2 package (version 1.18.1) (Love et al., 2014) was used to identify differential mRNA abundance levels. Analysis was carried out within R version 3.4.1 (https://www.r-project.org/). A p-adjust less than 0.05 was applied to the paired analysis results to distinguish genes with statistically significant changes in mRNA abundance. The p-adjust is the corrected p-value for multiple comparisons, reducing the number of false positives due to the number of tests. Alterations in mRNA abundance are expressed in log2 fold change, where a log2 fold change of 1 was used as a threshold for substantial change.

2.7.2 AS changes analysis

To detect differential alternative splicing in the RBPMS knockdown and overexpression, Dr. Miriam Llorian used rMATS v 3.2.5 (Shen et al., 2014), allowing for new splicing event discovery by applying flag novelSS 1. rMATS identifies splicing events from five main categories: skipped exons (SE), mutually exclusive exons (MXE), alternative $5'$ and $3'$ splice sites (A5SS and A3SS) and retained intron (RI). Furthermore, rMATS calculates the inclusion level and provides the statistical significance of each event identified.

We then applied an additional filter to the exon junctions and reads on target data to remove false positives generated by events with low read count. For that, all the identified events that showed a total of read counts in the replicates below 50 reads were discarded. In addition to that, an FDR less than 0.05 was imposed to the ASEs to distinguish statistically significant changes. Changes in ASEs were only considered substantial when displaying an inclusion level difference (ΔPSI) greater than $|10\%|$. For further analysis, unique IDs were created by combining the AS type, gene name and the chromosomal coordinates of the regulated exon and flanking exons separated by a ".".

Sashimi plots were generated by Dr. Miriam Llorian using rmats2sashimiplot (Gohr and Irimia, 2018). Sashimi plots allow visualization of the differentially regulated exons by displaying the RNAseq coverage and the reads mapping to the exon-exon junctions, which are represented by the arches. Moreover, they also show the PSI values calculated by the geometric mean of the PSI values from the triplicates.

ASEs identified in the RNAseq were experimentally validated by RT-PCR as described before in Section 2.6. In total twenty eight splicing events from knockdown and overexpression experiments were validated between me and Clare Gooding. Eventually, a Pearson correlation between RNAseq predicted and RT-PCR observed ΔPSIs was tested in RStudio (http://www.rstudio.com/).

Proportional Venn diagrams for visualization of different comparisons in this study were generated using BioVenn (Hulsen et al., 2008). Venn diagrams were produced to compare the overlap between genes whose mRNAs were regulated at the abundance level and also differentially spliced, and also to compare common AS events across RBPMS knockdown, overexpression and PAC1 dedifferentiation.

The experimental design of the RNAseq experiments, sample preparation and final data analysis are summarized in Figure 2.1.

2.7.3 Aorta Tissue dedifferentiation dataset

Previously in the laboratory, Adrian Buckroyd and Clare Gooding started the investigation of the AS in VSMC dedifferentiation. In order to do that, they isolated rat aortas from 8-12 weeks old Wistar rats. Adventitia was cleaned away by treating tissue samples with 3 mg/mL collagenase (Sigma C-0130) for 30 min at 37°C. Tissue samples were then treated to achieve dispersed single cells. Briefly, tissue pieces were treated with 5 ml 1 mg/mL elastase (Worthington Biochemical Corporation LS002292) for 30 min at 37°C followed by a further 1-2 h treatment with 5 mL collagenase (3 mg/ml). Dispersed single cells were washed, counted and $4x10^5$ cells transferred to 1 mL of M199 media containing 10% FBS, 2 mM Glutamine and 100 U/mL Penicillin-Streptomycin. SMCs were then further cultured to promote the proliferative state. For that, cells were maintained in DMEM media containing Glutamax and 10% FBS and passaged 1:2. The dedifferentiation of rat aorta VSMC is illustrated in Figure 2.2. Total RNA from three replicas of the following stages were extracted: aorta tissue (T), enzymatically dispersed single cultured SMCs (SC), passage 0 (P0) and passage 9 (P9). Each replica is a result of pooling five rat aorta dissections. Extraction was carried out using TRI-reagent (Sigma) and mRNA libraries prepared using Ribozero and TrueSeq kits. Sequencing was performed on a HiSeq2000 in a paired end mode at the Genomic Core

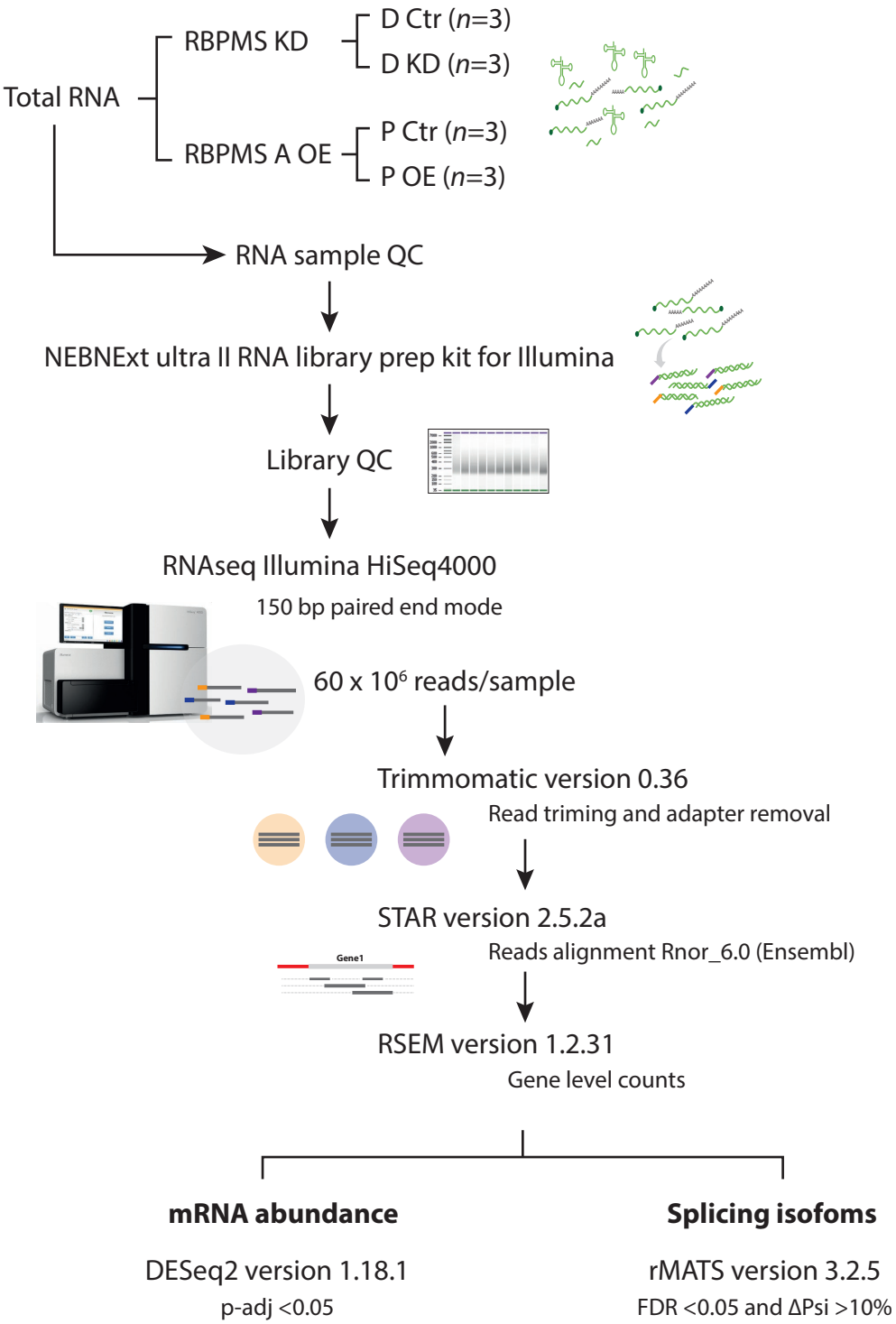


Fig. 2.1 RNAseq experiment and analysis workflow.

Facility at the EMBL Heidelberg. Dr. Miriam Llorian conducted the RNAseq analysis for mRNA abundance and alternative splicing changes using DEseq2 and rMATS as previously described in this section.

This tissue dataset was then used as a source of genes transcriptionally as well as AS regulated during aorta tissue dedifferentiation. As a note, in most of the comparisons with PAC1 experiments, the comparison between the most differentiated and most proliferative VSMC samples, T and P9 respectively, was used.

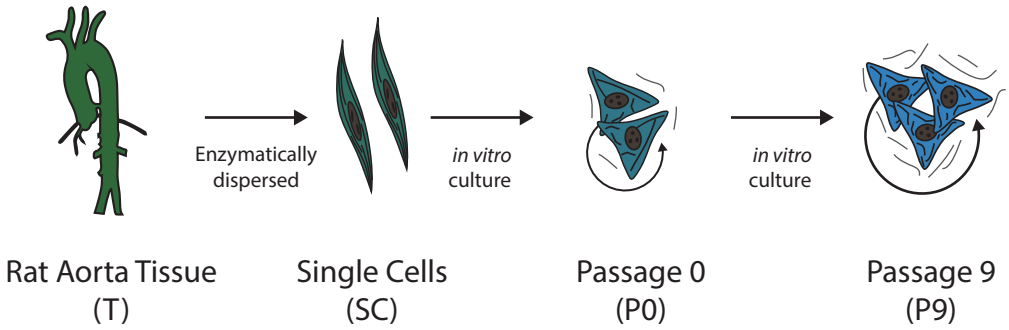


Fig. 2.2 Aorta dedifferentiation schematic.

2.8 RBPMS motif enrichment analysis

To identify RBPMS recognition elements around regulated skipped/cassette exons, the toolkit Matt (Gohr and Irimia, 2018) was used. Motif enrichment analyses were performed by Dr Aishwarya Jacob in the laboratory. Enrichment of optimal RBPMS binding sites, CAC-$N_{1\text{-}12}$-CAC (Farazi et al., 2014), were carried out for aorta tissue and PAC1 dedifferentiation and RBPMS depletion and overexpression data-sets. During the analysis ASEs were separated into three groups: 1) unregulated, 2) downregulated and 3) upregulated. Events were considered unregulated when showing an FDR between 0.1 and 1 and a $|\Delta PSI|$ lower than 5%. Significantly regulated events (FDR<0.05) were separated into downregulated, $\Delta PSI < -10\%$, and upregulated, $\Delta PSI > 10\%$. Intronic sequence analysis was performed by using 250 bp upstream and 250 bp downstream of the regulated skipped exon. For the motif analysis, Matt toolkit divides the exon and introns into three segments that are individually analyzed for motif fold enrichment and statistical significance. For the RBPMS maps, also generated using Matt, a sliding window of 31 was applied to the motif enrichment analysis that extended to the flanking

exons and far intronic sequences. The regions used in the motif analysis are further described in the figures.

2.9 Gene ontology analysis

Gene ontology (GO) terms enriched in the differentially abundant and spliced genes were obtained using Gorilla (Eden et al., 2009). A background list of all the genes expressed in either the PAC1 cells or aorta tissue was generated by filtering out genes whose expression was below 1 TPM (transcripts per million) in both conditions within the comparison. For gene expression, genes with statistical significant changes in mRNA abundance were divided into differentially upregulated or downregulated and the enrichment analysis individually carried for each list. For the differentially spliced genes, GO analysis was restricted to the SE category. Moreover, only genes containing AS events with $|\Delta PSI|$ greater than 10% were used in the analysis. Enriched terms were then sorted by p-value and only the top five significant GO terms from each category (biological process, cell component and molecular function) are shown in the results.

2.10 PPI analysis

To get insights into the potential protein-protein network affected by RBPMS splicing control, a PPI analysis was carried out using STRING version 10.5 (Szklarczyk et al., 2017). The PPI network was constructed focusing on core functions of RBPMS. Therefore, target events were restricted to events that were co-regulated between PAC1 dedifferentiation and RBPMS knockdown in combination with events commonly spliced in aorta tissue dedifferentiation and RBPMS overexpression. Furthermore, a $|\Delta PSI|$ greater than 10% was again applied. In view of its better curation, the human database was chosen for the analysis. Default parameters were also adjusted as followed: confidence as the meaning of the network edges, experiments and databases as the only sources of interaction and the minimum interaction score as high confidence (0.700). Functional GO enrichments were also retrieved from STRING analysis, although the whole human genome was used for the statistical enrichment in this case.

2.11 Recombinant protein

2.11.1 Induction of protein expression

Recombinant RBPMS-A and RBPMS-B containing a T7 and a 3xFLAG tag at the N-terminus and a His_6 tag at the C-terminus were expressed in *E. coli* (Figure 2.3). pET21d vectors containing either RBPMS-A or RBPMS-B were obtained as described in Section 2.2. BL21 competent *E. coli* cells were then transformed and a single colony was inoculated in 400 mL 2x yeast extract-tryptone (YT). At an OD of 0.7, 1 mM IPTG was added to induce expression of recombinant protein at 37°C for 3 h. Cells were resuspended in lysis buffer (50 mM $NaPO_4$ pH8, 200 mM KCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 0.05% NP-40, 5% glycerol and 1 mM PMSF) and lysed using a French Press. Lysates were centrifuged at 8.000 rpm for 10 min at 4°C to collect soluble and insoluble fractions. An SDS-PAGE gel was run for analysis of the soluble and insoluble fractions alongside uninduced and induced samples previously collected (Figure 2.3).

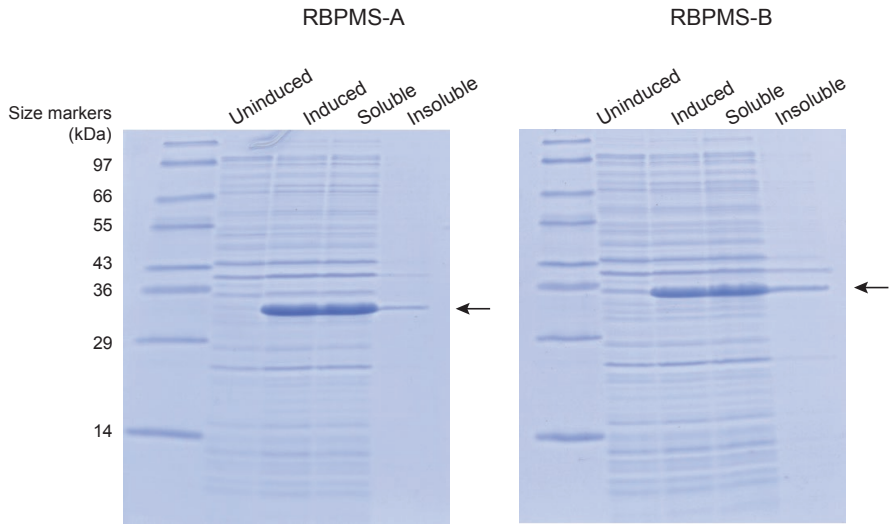


Fig. 2.3 RBPMS protein expression. Induction of RBPMS-A and RBPMS-B containing an N-terminal T7 and FLAG tags and a C-terminal His_6. Arrows indicate induced RBPMS isoforms.

2.11.2 Purification of RBPMS

Soluble recombinant RBPMS-A was then first purified using Blue Sepharose 6 and finally a HisTrap HP column in the AKTA protein purification system. Recombinant

RBPMS-B was purified only through the latter column, given the low affinity to the first purification column. After that, recombinant proteins were then dialysed overnight in buffer E (20 mM Hepes pH7.9, 20% glycerol, 0.1 M KCl, 0.2 mM EDTA and 0.5 mM DTT). Purity and integrity of the recombinant proteins were assessed by SDS-PAGE polyacrylamide electrophoresis. Furthermore, amino acid sequence was confirmed by western-blot using RBPMS specific antibodies and also by mass mapping by mass spectrometry. Proteins were quantified by BSA titration and NanoDrop spectrometer (Thermo Scientific).

2.12 *In vitro* transcription and binding assays

2.12.1 *In vitro* transcription

RBPMS direct binding to target RNAs was addressed by using two *in vitro* binding assays: electrophoretic mobility shift assay (EMSA) and UV-crosslinking. For both assays *in vitro* transcribed RNAs were produced from the constructs described in Section 2.2. Briefly, RNAs corresponding to the intronic regions containing CACs downstream of the regulated *FlnB* exon H1 and *Myocd* exon 2a and upstream of the *Tpm1* exon 3 were transcribed *in vitro* by SP6 RNA polymerase. The gel purified linearized DNA constructs containing the promoter sequence for SP6 RNA polymerase were used as templates. Each transcription reaction used 1 μg of DNA. For *in vitro* binding assays, high-specific activity protocol was performed. The reaction containing 2 μL of 5x transcription buffer (30 mM MgCl2, 10 mM spermidine, 0.2 M Tris-HCl, pH 7.5), 0.5 μL RNase inhibitor (Dundee Cell), 0.5 μL 10 mM RNA cap analogue (Promega), 1 μL water, 1 μ RNA polymerase and lastly 2 μL of 0.37 MBq/μL [α-^{32}P]UTP was then added to the transcription reaction for incorporation in to the RNAs. Reactions were incubated at 37°C for an hour and diluted in water in a 1:5 dilution. 1 μL of the dilution was separated for total count quantification. Finally, RNAs were purified by phenol extraction and through a G-50 column (ProbeQuant) by centrifugation at 3.000 rpm for 2 min to remove unincorporated nucleotides. Lastly, 1 μL of the final purified RNAs and the previously reserved 1 μL total sample were quantified using a Beckman LS 3801 scintillation counter. The [incorporated RNA counts x final volume]/[total RNA counts x 50 μL] could then be used to calculate the percentage yield. Amount of RNA transcribed is finally calculated as [%yield/100] x [mass of the nucleotide x number of nucleotides x UTP concentration x volume]. In vitro transcribe RNAs were then diluted to 20 fmol/μL.

2.12.2 EMSA

EMSA or band shifts were performed to detect binding between *in vitro* transcribed RBPMS-splicing target RNAs and RBPMS. Firstly, binding reactions were achieved by incubating the recombinant protein in a serial dilution (0, 0.125, 0.5 and 2 μM) with 10 fmol of *in vitro* transcribed RNAs in binding buffer (10 mM Hepes pH7.2-7.9, 3 mM $MgCl_2$, 5% glycerol, 1 mM DTT and 40 mM KCl) for 25 min at 30°C. 1μL of 50% glycerol was added to the reactions that were then loaded onto a 4% 60:1 acrylamide and Bis ratio nondenaturing gel. "Conventional" gels were initially used (6% 30:1 acrylamide and Bis ratio), but RBPMS-RNA complexes did not go through the gel and remained retained in the well. Therefore, a different Bis/acrylamide ratio as well as percentage were optimized for the band shift assays. Gels were then dried and Fuji Medical X-ray film used for detection. Films were then developed.

2.12.3 UV Crosslink

For UV-crosslinking, the same binding reaction for EMSA was performed followed by UV-crosslink on ice in a Stratalinker with 1920 mJ. Next, an RNase mix, containing RNase A and T1, was added and reactions further incubated for 10 min at 37°C. Finally, 4X SDS sample buffer was added to the binding reactions which were then heated at 80°C for 5 min prior to loading on a 20% denaturing polyacrylamide gel. Gels were then dried and Fuji Medical X-ray film used for detection. Films were then developed.

2.13 Statistical analysis

Statistical tests applied in each experiment are explained in its corresponding method section and it is also described in the figure legends. In all the experiments, p-value is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Data analysis and visualization was carried out using RStudio (http://www.rstudio.com/).

2.14 Data availability

All mRNAseq data from this study have been deposited in NCBI Gene Expression Omnibus (GEO) repository under GEO accession GSE127800, in which GSE127799 represents the RBPMS knockdown and overexpression in rat PAC1 cells and GSE127794 the rat aorta tissue dedifferentiation in culture.

Chapter 3

RBPMS: a potential SMC master regulator

3.1 Introduction

3.1.1 Splicing master regulators

For a long time, cell-identity has been studied under the transcription perspective, which led to the identification of a small number of transcription factors so called master transcription factors (Young, 2011). These factors determine cell states by initiating the gene expression program of a particular cell type. Interestingly, the existence of similar regulators at the post-transcriptional level has recently been proposed. Such "master splicing regulators" would consist of a subset of splicing factors able to reshape and maintain a tissue transcriptome upon external cues during development (Jangi and Sharp, 2014). Indeed, AS contributes to a layer of a cell specific transcriptome by generating protein isoforms and also quantitatively regulating transcripts via NMD. Therefore, the highly interconnected splicing regulated genes form a robust cell specific network that retains its capability to respond to environmental stimuli (Jangi and Sharp, 2014).

For Jangi and Sharp, some of the properties expected for a master regulator would be:

1. Requirement for differentiation, specification and/or maintenance of a cell-type
2. Expression in a tissue-specific manner
3. Wide dynamic range of activity not limited by auto-regulation

4. Highly regulated only responding to external signals, in other words not regulated from within the splicing network itself

5. Likely to control transcription factors and other post transcriptional regulators

6. Regulation of functionally related direct and indirect targets that are important for cell phenotype and identity

Taken these hallmarks of master splicing regulators, manipulation of their levels such as by depletion or misexpression would result in either incomplete differentiation or promotion of the differentiation towards the cell-type in which that master regulator is normally expressed (Jangi and Sharp, 2014). Even though these characteristics distinguish master regulators from other splicing factors, identification of these regulators in a cell type is still very difficult.

Master regulators of splicing could be identified by either their response to particular cell type signal or by the tissue specific transcriptional control (Jangi and Sharp, 2014). Due to the difficulties faced by the former approach (Heyd and Lynch, 2011), the latter seems to be the best strategy for the identification of cell-type master regulators. In fact, Jangi and Sharp suggested key splicing factors of a tissue are likely to be activated by super-enhancers (Figure 3.1). Thereby, master regulators could potentially be identified via the association of their genes with super-enhancers.

3.1.2 Super-enhancer associated RBPs

Different from normal transcriptional enhancers, super-enhancer regions consist of clusters of enhancers highly occupied by transcription factors, driving high-level of expression of their associated genes (Hnisz et al., 2013). Tissue-specific super-enhancers drive the expression of genes important for cell identity and function as revealed by a catalog of human tissue super-enhancers generated in (Hnisz et al., 2013). Super-enhancer sequences are distinguished by the enrichment for transcriptional machinery components, histone modifications marks such as methylation and acetylation (H3K4me1 and H3K27ac) and lastly the presence of conserved binding sites for transcription factors (Shin, 2018). Therefore, it is plausible that RBPs activated by super-enhancers could be critical for the reshaping of a cell-type transcriptome, acting as master splicing regulators, as suggested in Jangi and Sharp (2014).

RBM24 and RBM38 are examples of post-transcriptional master regulators of myogenesis whose gene expression is driven by super-enhancers (Jangi and Sharp, 2014). *Rbm24* and its paralog, *Rbm38*, are marked by super-enhancers in adult skeletal

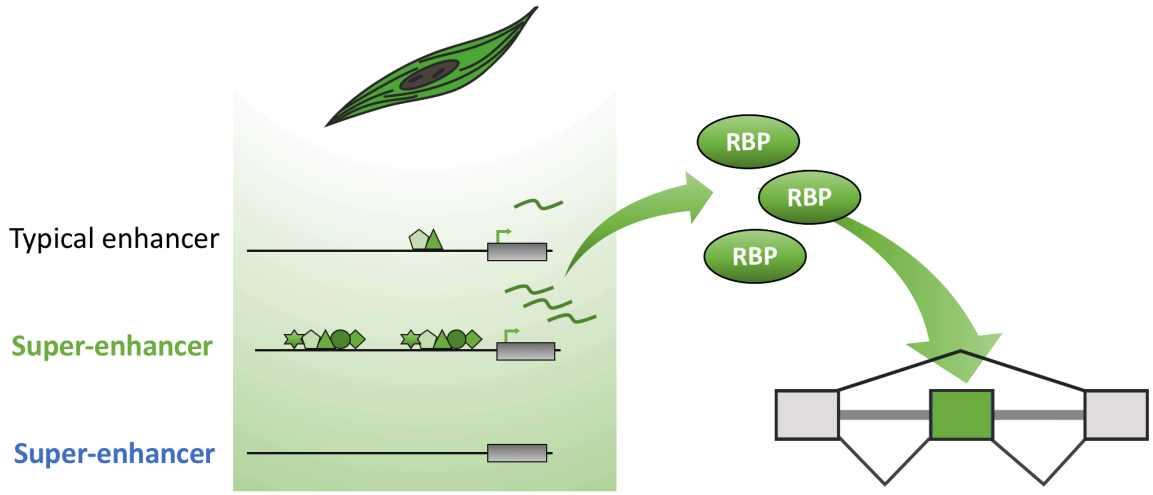


Fig. 3.1 Super-enhancer associated RBPs could act as master splicing regulators. Illustration of a cell type specific super-enhancer associated with the activation of a master splicing regulator that can then reshape the cell transcriptome, as suggested in Jangi and Sharp (2014).

muscle and have been implicated in myogenesis by post-transcriptionally controlling p21 (Jin et al., 2010; Miyamoto et al., 2009). Both proteins are known to be activated by the master transcription factor, MyoD, early in the differentiation, which leads to the stabilization of myogenin and p21, an important myogenesis factor and a cell-cycle key regulator respectively (Jin et al., 2010; Miyamoto et al., 2009). Moreover, these two RBPs can build a robust skeletal-muscle splicing network by co-regulating events with another splicing regulator, Rbfox1 (Zhang et al., 2008).

Finally, it has also been shown that super-enhancers are enriched for disease-associated variations, expanding the implications of super-enhancers from healthy tissues to also disease states. Not surprisingly, super-enhacers have more recently also been involved in tumorigenesis, in which acquisition of super-enhancers were observed for oncogenic genes (Hnisz et al., 2015). Therefore, super-enhancers have also been exploited as targets for disease treatment and diagnosis, by small molecule inhibitors and potential gene therapy (Shin, 2018).

In this chapter we set to find RBPs that could act as master splicing regulators of the SMCs via their association with super-enhancers, thereby identifying regulators that could potentially play a critical role in establishing or maintaining the differentiated splicing program of these cells.

3.2 Results

3.2.1 RBP genes associated with SM super-enhancers

To identify potential master splicing regulators of SM tissues, publicly available datasets of super-enhancer associated genes across different human tissues (Hnisz et al., 2013) and the curated list of human RBPs containing 1542 proteins (Gerstberger et al., 2014) were used. Datasets of super-enhancer associated genes from aorta (1143 genes), bladder (412 genes) and stomach SM (974 genes) were selected as SMC rich tissues as described in Chapter 2. In addition to that, skeletal muscle (1078 genes) was chosen as an outlier tissue dataset. Firstly, the RBP genes associated with super-enhancers in each tissue were identified. 67, 36, 70 and 98 RBP genes were found associated with super-enhancers in aorta, bladder, stomach SM and skeletal muscle respectively. Next, RBP genes associated with SM tissues but not skeletal muscle were found by comparing the lists as shown by the Venn diagram (Figure 3.2). Only nine RBP genes met this criteria: CPSF4L, DHX16, IMP3, LSM4, RBPMS, RPL10A, RPL11, SETD1B and TNRC6C. Furthermore, using a set of super-enhancer associated genes

in the dbSUPER database (Khan and Zhang, 2016), two RBP candidates were found in SM tissues, of which only RBPMS was shared with the original nine candidates (Figure 3.2).

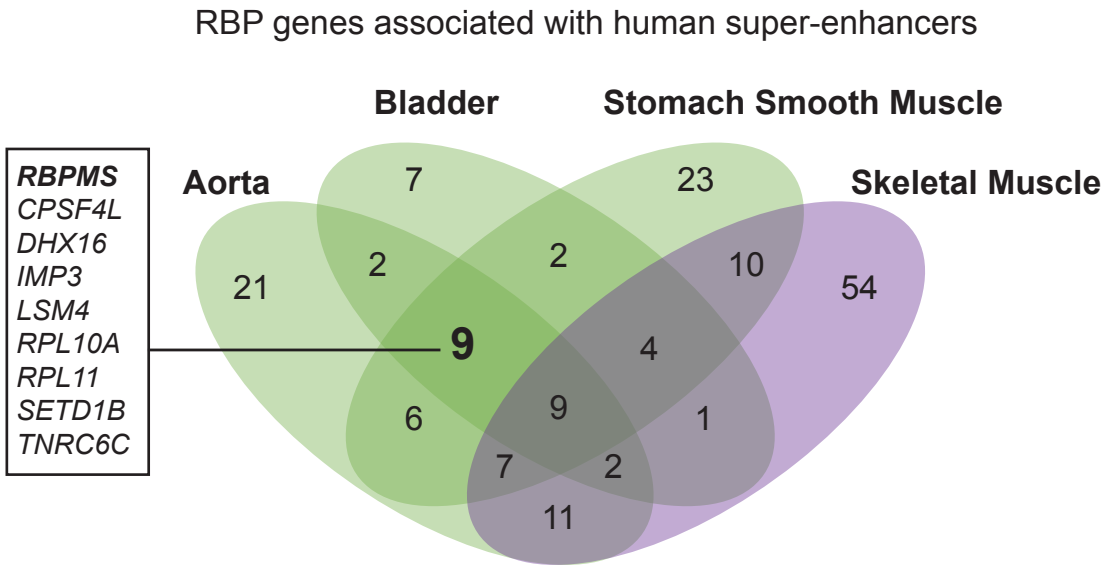


Fig. 3.2 Nine RBP genes were found associated with human SM tissue super-enhancers. Venn diagram comparing the RBP genes associated with human super-enhancers from three SM tissues (aorta, bladder and stomach - all in green) and one outlier tissue, skeletal muscle (purple). On the left, RBP genes found specifically associated with SM super-enhancers.

The mRNA levels of these RBP genes were then investigated in the rat aorta tissue dedifferentiation dataset (samples generated in the laboratory by Adrian Buckroyd and Clare Gooding and RNAseq analyzed by Miriam Llorian - Chapter 2.7.3). mRNA abundance levels were obtained from rat aorta tissue (T), enzymatically dispersed SMCs (SC), and from SC grown in cell culture to promote dedifferentiation and harvested at passage 0 (P0) and passage 9 (P9). Expression data was retrieved for all the candidate regulators, but *Cpsf4l*, which was not found in the RNAseq analysis (Figure 3.3). From the eight RBP genes only *Rbpms*, *Setd1b* and *Tnrc6c* showed a significant gradual decrease in their mRNA abundance during dedifferentiation. Furthermore, in terms of the absolute expression levels, measured in transcripts per million (TPM), apart from the ribosomal protein coding genes, *Rpl10a* and *Rpl11*, *Rbpms* was the only gene that showed considerable expression level in aorta tissue ($\sim$ 50 TPM), which was around five-fold greater than the other candidates ($\sim$10 TPM).

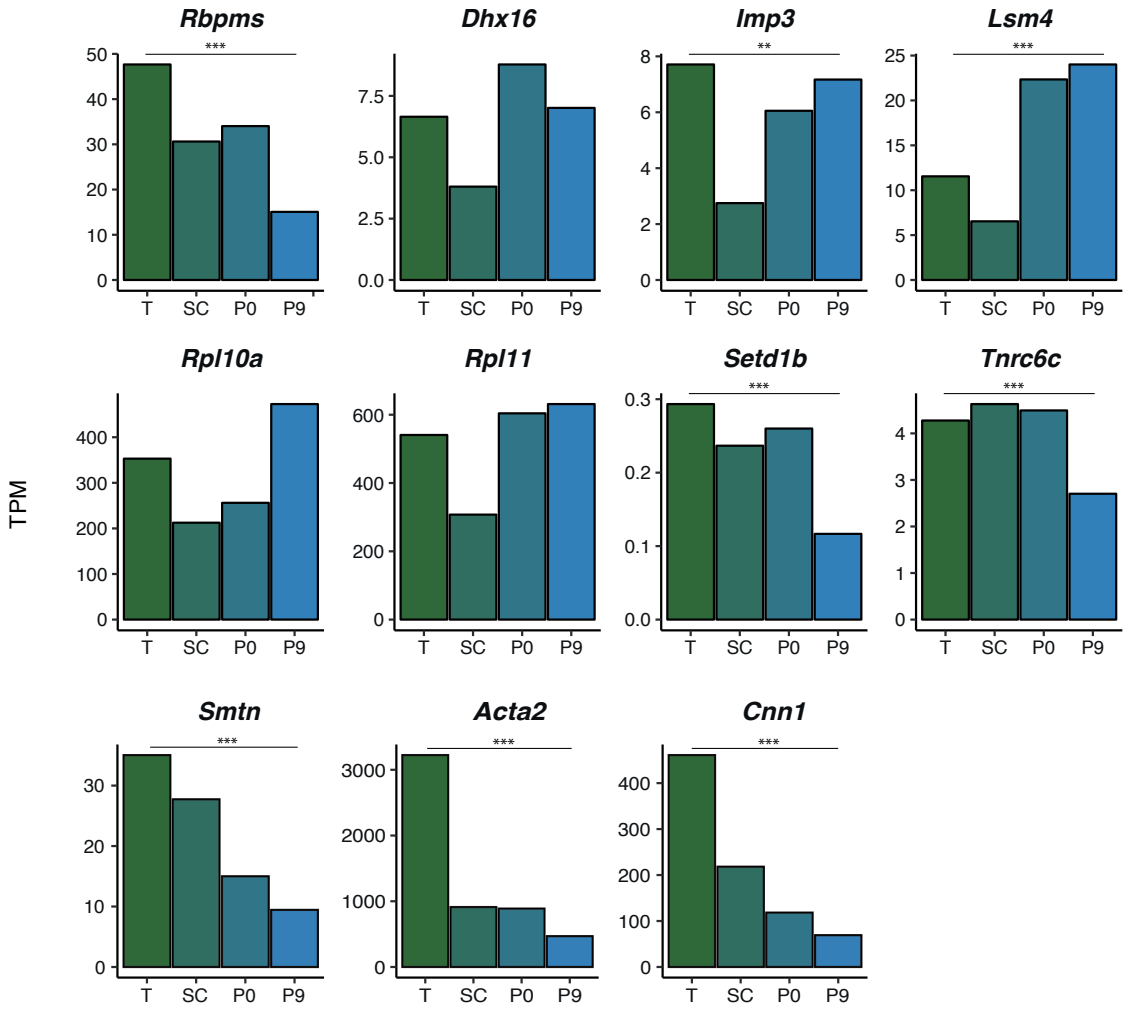


Fig. 3.3 mRNA expression levels of potential SM master regulators in rat aorta dedifferentiation. mRNA expression of the nine RBP identified associated with SM super-enhancers was assessed during rat aorta dedifferentiation. Expression in transcripts per million (TPM) is shown for aorta tissue (T), enzymatically dispersed SMCs (SC), passage 0 (P0) and passage 9 (P9). *Smtn*, *Acta2* and *Cnn1* are shown as SMC markers. Statistical significance calculated by DESeq2 between T and P9 is displayed as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. *Cpsf4l* was not found to be expressed in the rat VSMCs.

Rbpms expression during rat aorta dedifferentiation was then compared to its paralog *Rbpms2* levels (Figure 3.4). *Rbpms* mRNA abundance decreased ∼4-fold from tissue to passage 9 whereas *Rbpms2* showed a greater fold change in its mRNA level, ∼8 fold lower in P9. However, *Rbpms2* mRNA absolute level (∼4 TPM) was approximately ten times lower than *Rbpms* expression (Figure 3.4). Moreover, changes in *Acta2*, a SMC transcriptional marker, are consistent with the rat aorta dedifferentiation. Changes in the expression of other RBPs known to regulate AS in SMC, e.g. MBNL1, PTBP1 and QKI, were either small or increased in aorta proliferative state (P9) (Figure 3.4). In addition to that, all these RBPs were expressed at similar levels at their maximum expression level (40 to 80 TPM range).

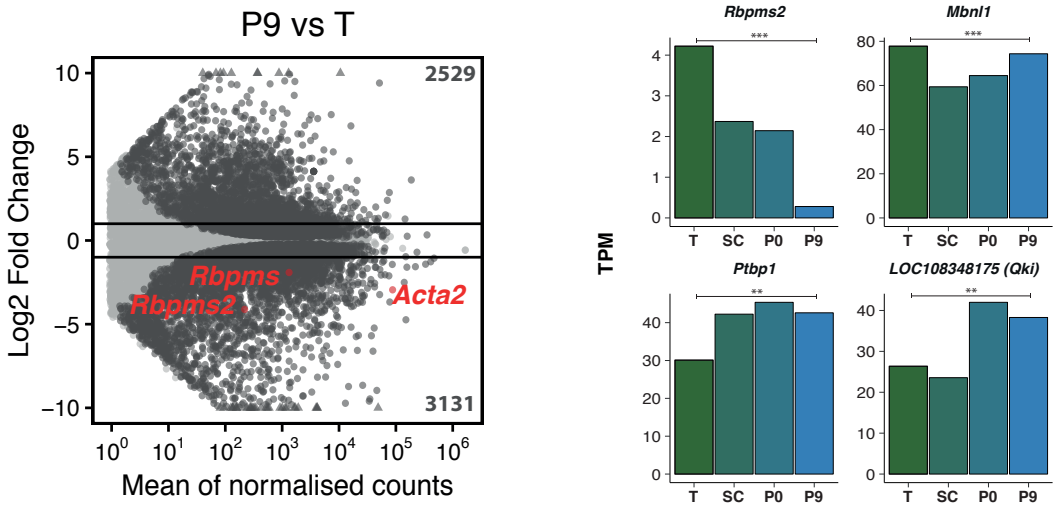


Fig. 3.4 *Rbpms2* is less expressed in VSMC. Left, MA plot for the alterations in mRNA abundance between rat aorta tissue and passage 9 (P9 vs T). Genes with significant change (padj<0.05) in their mRNA abundance are coloured in dark gray. Genes with p-adj values ≥0.05 are shown in light gray. *Rbpms*, *Rbpms2* and *Acta2*, a SMC marker of differentiation, are highlighted in red. Horizontal lines indicate log2 fold change 1 and -1. Numbers found at top and bottom right indicate upregulated and downregulated genes, respectively. **Right**, mRNA expression level of *Rbpms2* and other RBPs (*Mbnl1*, *Ptbp1* and *Qki*) during rat aorta dedifferentiation are shown in transcripts per million (TPM). Aorta tissue (T), enzymatically dispersed SMCs (SC), passage 0 (P0) and passage 9 (P9). Statistical significance calculated by DESeq2 between T and P9 is displayed as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

RBPMS transcript expression was further examined across different human tissues in the Genotype-Tissue Expression (GTEx) project (Lonsdale et al., 2013) (Figure 3.5). RBPMS was highly expressed in SM tissues, as highlighted by the top ten human

tissues expressing RBPMS, which includes bladder and three different arteries (tibial, aorta and coronary).

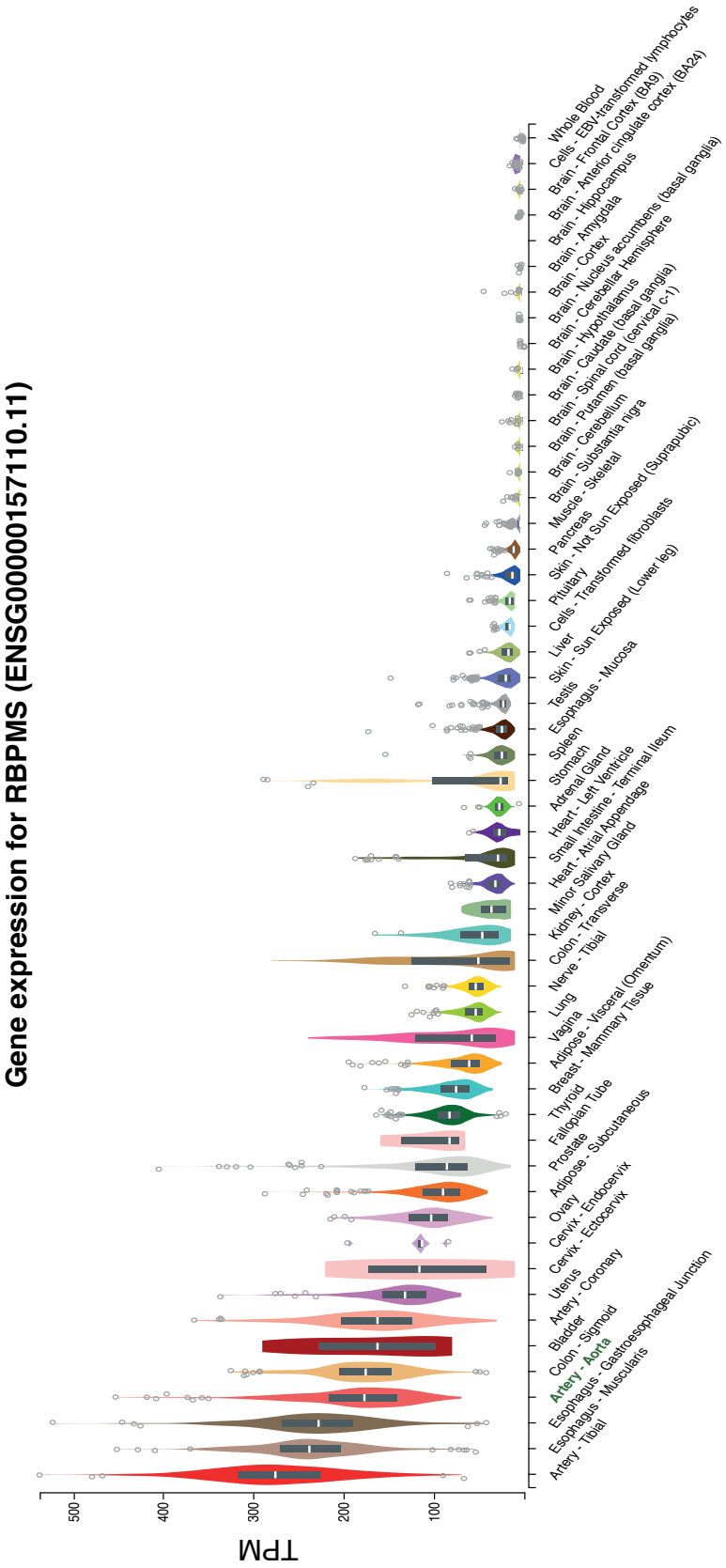


Fig. 3.5 RBPMS is highly expressed in human SM tissues. mRNA expression of RBPMS across different human tissues was obtained from GTEX (Lonsdale et al., 2013). Aorta tissue is highlighted in green font. Expression is shown in transcripts per million (TPM)

3.2.1.1 RBPMS in PAC1 cells

Rbpms expression was also assessed in the rat pulmonary arterial SM cell line, PAC1 (Rothman et al., 1992) (Figure 3.6). This cell line maintains differentiated properties in culture and can also be grown towards a more proliferative phenotype in culture as described in (Llorian et al., 2016). Therefore, PAC1 cells allow modeling of SMC phenotypic modulation in culture. In the PAC1 cells, qRT-PCR revealed that *Rbpms* mRNA expression was also downregulated in the proliferative state, ∼5 fold reduction compared to the differentiated PAC1 (Figure 3.6A). On the other hand, *Rbpms2* expression was not regulated during PAC1 dedifferentiation. PAC1 dedifferentiation was confirmed by the downregulation of SMC transcriptional markers (*Acta2*, *Cnn1* and *Smtn*). RBPMS protein levels were then assessed by western blotting, showing that RBPMS protein was reduced to below detection in the proliferative cells (Figure 3.6B). The SMC marker ACTA2 was also only detected in the differentiated PAC1 state. In agreement with the qRT-PCR and western blot, immunofluorescence confirmed RBPMS high expression in the differentiated PAC1 cells (Figure 3.6C). Furthermore, immunofluorescence disclosed RBPMS predominantly nuclear localization, consistent with its potential role as a splicing regulator.

Therefore, RBPMS association with SM super-enhancers, its high expression in human SM tissues, modulation of its expression during rat aorta and PAC1 dedifferentiation, together with its nuclear localization supported this RBP as a promising candidate for a master post-transcriptional regulator of SMCs.

3.2.2 Splicing regulation by RBPMS

To test the hypothesis that RBPMS regulates alternative splicing in SMC, rat *Rbpms* cDNAs were cloned from differentiated PAC1 cells (Chapter 2.2.7). In total, seven protein isoforms were identified and their exon arrangement are found in Figure 3.7. The differences in the protein sequence of isolated RBPMS isoforms and RBPMS2 are highlighted in the alignment in Figure 3.8. The two main isoforms identified were RBPMS A and B, determined by the inclusion of either exon 7 (alternative 3′ end) or exon 8. Consequently this ASE affects the C-terminus of the protein isoforms (Figure 3.9). Hence, the two bands observed in the western blot corresponded in size to RBPMS-A and B (Figure 3.6B). The other RBPMS minor isoforms, apart from *Rbpms-A.3*, are mainly generated by exclusion of exon 6. In that way, they all contain an intact RRM domain (encoded by exons 2-5) with protein variation mainly restricted to their C-terminus (Figure 3.7 and 3.8). In *Rbpms-A.3* transcript isoform, exons

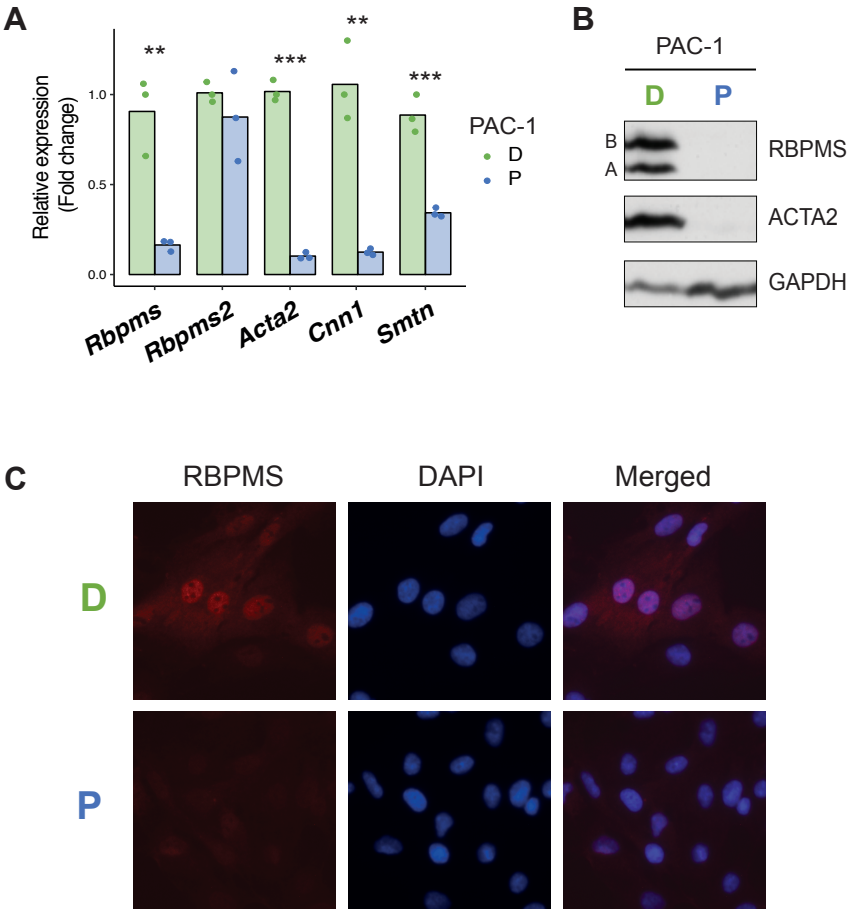


Fig. 3.6 RBPMS is strongly downregulated during PAC1 cells dedifferentiation. **A** Relative mRNA expression levels of *Rbpms* and *Rbpms2* in differentiated (green) and proliferative (blue) PAC1 cells quantitated by qRT-PCR. Expression levels of SMC differentiation markers, *Acta2*, *Cnn1* and *Smtn*, are also shown. Points in the bar graphs show data from individual samples from the triplicate and the mean is indicated with the bar. Statistical significance from a two tailed unpaired Student's t-test is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. **B** Western blot for RBPMS in rat PAC1 cells. ACTA2 was used as a SMC differentiation marker and GAPDH as a loading control. **C** Immunofluorescence in rat PAC1 cells for RBPMS. Middle panel show DAPI staining for nuclei. Differentiated and proliferative PAC1 cells are indicated as D and P, respectively.

2-4 are not included, encoding an RBPMS isoform lacking most of its RRM domain (Figure 3.7 and 3.8). On the other hand, only one isoform from *Rbpms2* was cloned from PAC1 cells. This single RBPMS2 isoform shares ∼70% identity with RBPMS-A.

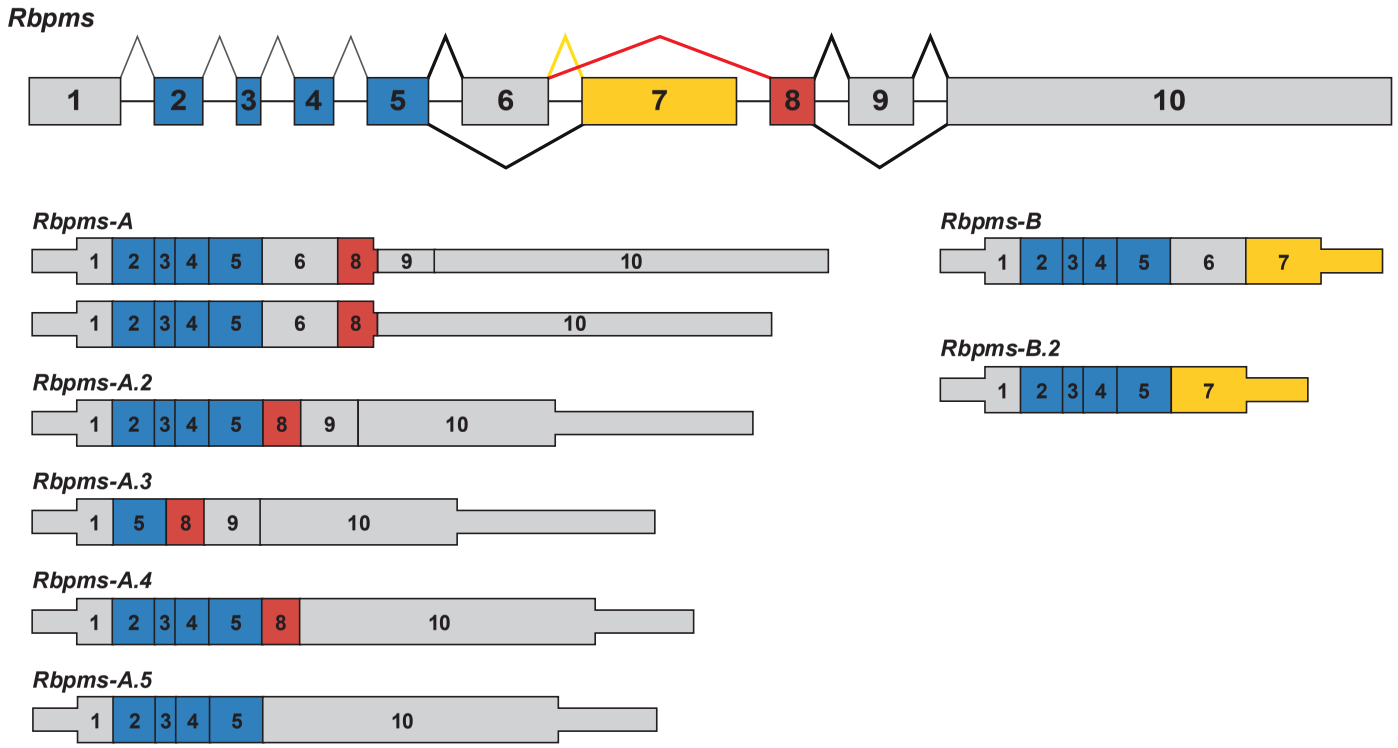


Fig. 3.7 *Rbpms* is expressed in multiple mRNA isoforms. Schematic of RBPMS transcript isoforms. **Top**, arrangement of *Rbpms* exons. **Bottom**, RBPMS mRNA isoforms. Introns are not shown in scale. Untranslated regions are represented by the narrow boxes and exons coding for the RNA Recognition Motif (RRM) are in blue. Exons 7 and 8, in yellow and red respectively, represent the alternative 3′ exons that determines RBPMS C-terminus. Arches indicate the exon junctions and the thick lines highlight alternative splicing events.

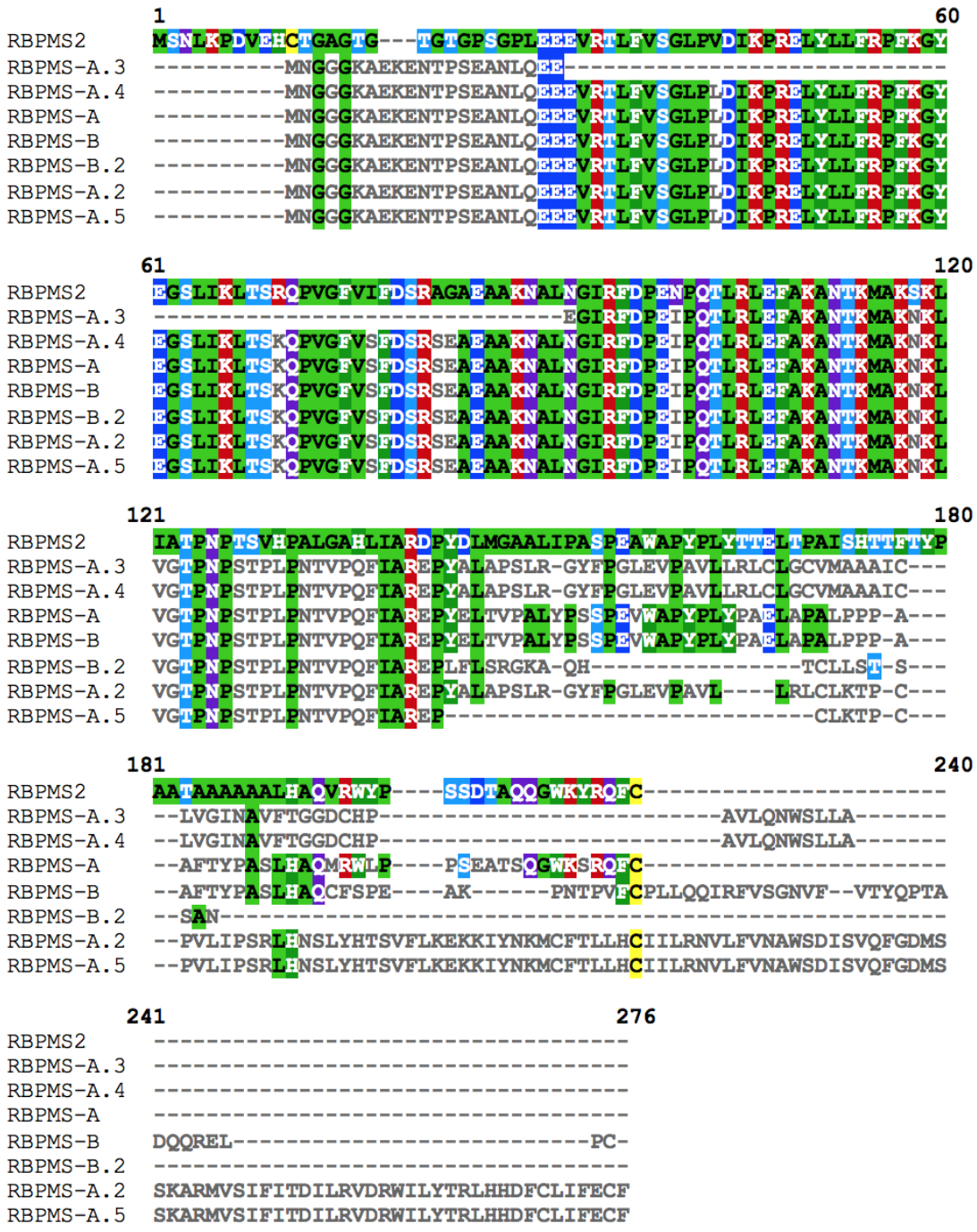


Fig. 3.8 Alignment of RBPMS protein isoforms and RBPMS2 identified in PAC1 cells Multiple alignment of rat RBPMS isoforms and RBPMS2 was generated using CLUSTAL O (1.2.4) and visualized with MView. Amino acids were colored using RBPMS2 as the reference sequence (bright green - hydrophobic, dark green - large hydrophobic, purple - polar, red - positive charge, dark blue - negative charge, bright blue - small alcohol, yellow - cysteine).

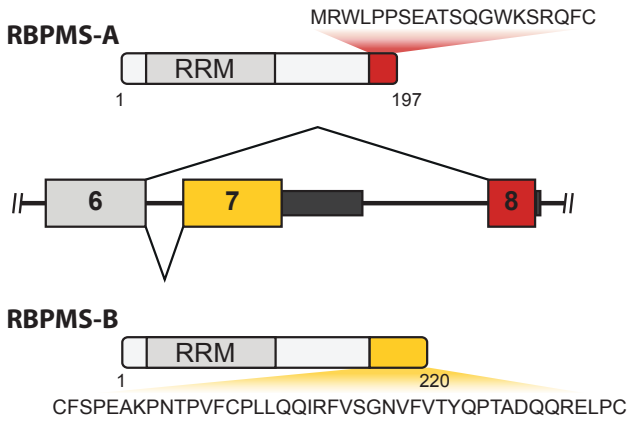


Fig. 3.9 RBPMS-A and RBPMS-B are the major isoforms in PAC1 cells Schematic of the two main RBPMS protein isoforms identified in rat differentiated PAC1 cells. The alternative 3′ end exon that gives rise to the RBPMS isoforms is shown in the middle. Amino acid sequences of the distinct C termini are highlighted.

Because RBPMS motifs had previously been resolved as tandem CACs separated by variable spacer length (1-12 nt) (Farazi et al., 2014), an RBPMS motif enrichment analysis was carried out for skipped exons differentially regulated between rat aorta tissue and passage 9 (Figure 3.10). Significant enrichment for RBPMS binding sites was observed upstream of and within exons more included in proliferative cells (∼1.4 fold) (Figure 3.10). Interestingly, this location of RBPMS motifs, if in agreement with a position-dependent alternative splicing activity shown for many other splicing regulators (Fu and Ares, 2014; Witten and Ule, 2011), suggests RBPMS as a repressor of proliferative exons. Thus, RBPMS binding sites around exons regulated during rat aorta dedifferentiation further supported a potential function of RBPMS as a splicing regulator in SMCs.

Consistent with the RBPMS motif enrichment analysis, SMC-specific splicing events previously characterized in the laboratory, such as the mutually exclusive pair of exons in *Tpm1* and *Actn1* (Gromak and Smith, 2002; Southby et al., 1999), contain conserved clusters of CACs (Figure 3.11). Conserved CACs were found around *Tpm1* exon 3 and upstream of *Actn1* NM exon (Figure 3.11). These two exons are repressed in differentiated SMC, being more included in proliferative SMCs. So, both *Actn1* and *Tpm1* exons were expected to be targets for repressive regulation by RBPMS.

To address whether RBPMS could regulate AS of SMC exons, RBPMS isoforms and paralog were tested upon regulation of the SMC minigene splicing reporters *Tpm1*

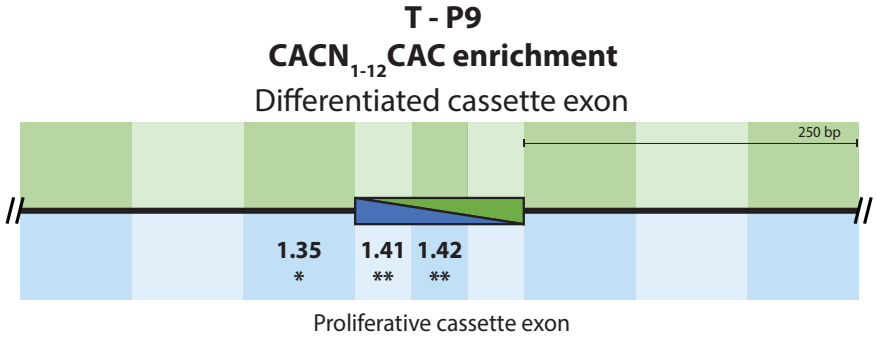


Fig. 3.10 RBPMS motifs are enriched upstream of exons less included in aorta tissue. Enrichment for optimal RBPMS motif, a pair of CAC separated by 1-12 nt (Farazi et al., 2014), around regulated skipped exons in the rat aorta tissue - passage 9 comparison (T - P9) was obtained by using *Matt* toolkit (Gohr and Irimia, 2018). Motif enrichment analysis for aorta differentiated (green) and proliferative (blue) exons. Numbers represent the motif fold enrichment and statistical significance is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

(Figure3.12A) and *Actn1* in PAC1 cells (Figure3.12B). Initially, all the RBPMS isoforms cloned from PAC1 cells were tested, but exon 6 skipped isoforms were either expressed at very low levels or not even detected. This was part of the evidence for dismissing some of these isoforms. Also the C-terminal amino acid sequences of these framshifted isoforms are less conserved than RBPMS-A and B, consistent with them not being functional protein-coding isoforms. Therefore all the overexpression experiments were carried out only using RBPMS-A, RBPMS-B and RBPMS2. In both experiments, PAC1 transfections were carried out in proliferative PAC1 cells, in which RBPMS is less expressed therefore its overexpression would lead to stronger effects. Indeed, both minigene reporters showed a starting splicing pattern of less differentiated SMC - nearly complete inclusion of the *Tpm1* exon 3 and the *Actn1* exon NM (Figures 3.12). RBPMS was sufficient to promote the differentiated splicing pattern, yet different splicing activity was observed between isoforms. RBPMS-A overexpression led to a switch of *Tpm1* and *Actn1* splicing isoforms, by reducing inclusion of exon 3 from 94.7% to 43.5% and exon NM from 100% to 21%. RBPMS2 showed similar activity to RBPMS-A but switching splicing to a less extent. Strikingly, RBPMS-B was the least active protein causing minor changes in splicing of *Tpm1* reporter, ∼10% skipping of exon 3 compared to Venus and no changes in *Actn1* exon NM splicing. RBPMS-B lower activity in repression of *Tpm1* exon 3 could be due to its lower protein expression compared to RBPMS-A, as indicated by the western blotting. On the other hand,

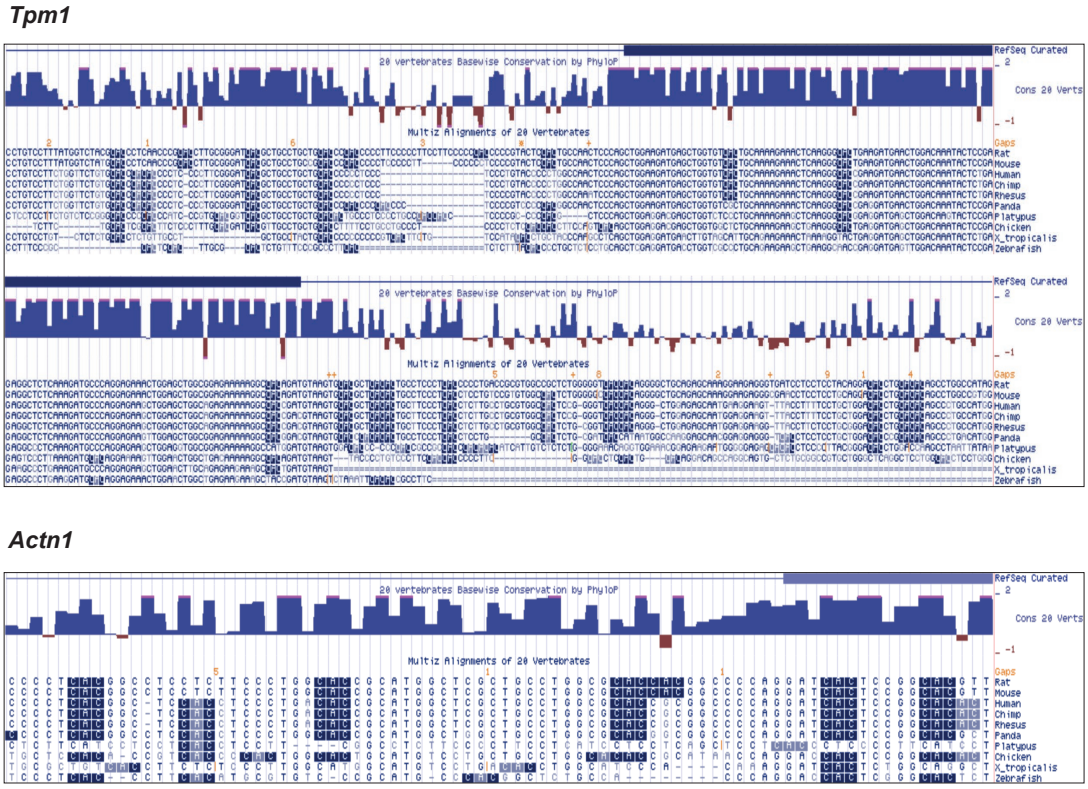


Fig. 3.11 SMC splicing markers contain conserved CAC motifs. Screenshots from USCS genome browser. **A** Conserved CAC motifs upstream, top, and downstream, bottom, of *Tpm1* exon 3 are highlighted in blue. Top dark blue box correspond to *Tpm1* exon 3. **B** CAC motifs conserved upstream of *Actn1* exon NM is shown in blue. Top purple box indicates *Actn1* exon NM. Basewise conservation track and multiple alignment were generated by PhyloP and Multiz Alignments in the UCSC genome browser. Chr coordinates from *rn6*: chr8:72,840,094-72,840,283; chr8:72,839,906-72,840,095; chr6:103,379,884-103,379,963.

Actn1 exon SM was only included upon RBPMS-A overexpression despite RBPMS-B being more expressed than RBPMS-A in this experiment. Thus, at least in the case of *Actn1* splicing, RBPMS isoforms showed differential activity not associated with its protein expression levels.

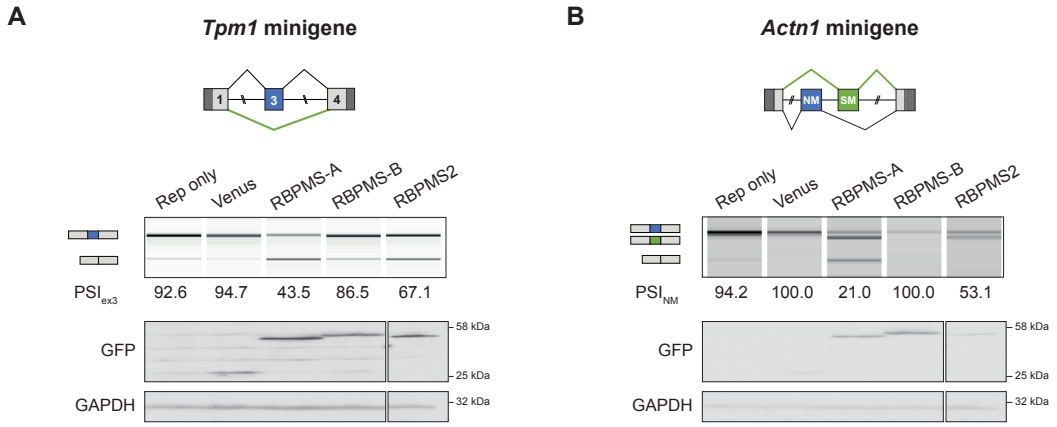


Fig. 3.12 RBPMS promotes the SM splicing pattern of *Tpm1* and *Actn1* in PAC1 cells. A Splicing of *Tpm1* minigene reporter was tested upon overexpression of Venus tagged RBPMS isoforms, A and B, and RBPMS2 in PAC1 cells. Reporter only and Venus controls were also tested. Top, RT-PCRs to detect exon 3 included and excluded isoforms, larger and smaller products, respectively. Moreover, schematics on the left indicates the RNA isoforms. Percentage of spliced in (PSI) values for exon 3 are shown below the gel image. Bottom, western blot anti-GFP for detection of the expression of Venus tagged proteins. GAPDH was used as a loading control. **B** Splicing of the *Actn1* minigene reporter was also tested upon overexpression of Venus tagged RBPMS isoforms A and B, and RBPMS2 in PAC1 cells as similarly described for the *Tpm1* reporter. For the *Actn1* minigene reporter, the PSI of the NM exon is shown. Top, RT-PCR for the different splicing isoforms and bottom, western blot to verify overexpression. In the schematics for the RNA splicing isoforms, green indicates differentiated exons and blue proliferative exons. A minus RT and a no template control were also carried out for every RT-PCR (data not shown). Protein size markers are also indicated on the right of the western blots. Both experiments are n =1, so no statistical significance test was performed for any of the experiments.

Due to the difficulties in achieving reasonable transfection efficiency in PAC1 cells, RBPMS ability to switch AS was tested in the human cell line HEK293, which is generally easier to transfect (Figures 3.13). Rat RBPMS-A, RBPMS-B and RBPMS2 were transiently overexpressed in HEK293 cells and changes in endogenous AS assessed. As expected from non-smooth muscle cells, *TPM1* and *ACTN1* smooth muscle specific

exons were not included at all in the Venus controls (Figures 3.13). Transfection of RBPMS-A, RBPMS-B and RBPMS2 caused a substantial shift towards *TPM1* exon 2 (∼37% average inclusion), which is more included in differentiated SMC. On the other hand, RBPMS effects on *ACTN1* splicing in HEK293 cells were similar to the regulation of the *Actn1* minigene reporter in PAC1 transfections. RBPMS isoforms showed differential activity, where RBPMS A and 2 were the most active proteins despite being less expressed then RBPMS-B. RBPMS-A and RBPMS2 produced more *ACTN1* transcripts with the NM exon excluded in HEK293 cells, ∼60% (Figure 3.13).

In addition to the proliferative exons, two new activation events were also tested, *FLNB* exon H1 and *MPRIP* exon 23. These splicing targets were identified by re-analysis of RBPMS overexpression in HEK293 RNAseq data from Farazi et al. (2014). This reanalysis of the RNAseq data was performed by Dr. Miriam Llorian using rMATS instead of DEXSeq as in the original study. Validation of these differentiated events surprisingly showed that all the proteins were able to activate splicing of both exons. RBPMS promotion of inclusion was clear in the HEK293 cells (Figure 3.13) in which the basal inclusion levels of *FLNB* and *MPRIP* in the controls were low (∼30% and 7% respectively). RBPMS overexpression in HEK293 cells increased the inclusion of *FLNB* exon H1 by ∼10 folds (from 5% up to 57%). Exon 23 was included in the majority of *MPRIP* transcripts upon overexpression of RBPMS proteins, approximately 70%. Therefore, these data indicate that RBPMS can regulate splicing by activating SM exons and repressing NM exons and that RBPMS is also sufficient to promote the SM differentiated splicing in non-smooth muscle cells. Curiously, RBPMS-A was more active than isoform B, despite higher expression of RBPMS-B than RBPMS-A in some experiments. In that way, this suggests that the differential activity between isoforms results from their distinct protein sequences.

The human cell line, Hela, was also tested upon RBPMS overexpression (Figure 3.14). In HeLa cells, endogenous *TPM1* was not detected, so RT-PCRs are not shown. Likewise HEK293 and PAC1 transfections, RBPMS was able to regulate splicing of *ACTN1*, *FLNB* and *MPRIP*. However, the effects upon the activation events, *FLNB* and *MPRIP* were less impressive, mainly because the starting splicing pattern is already most of the way towards inclusion. Thus, RBPMS could not promote the inclusion pattern much more, yet significant changes were observed. Once more, differential splicing activity in the regulation of *ACTN1* was observed, being RBPMS-B less active, even though its protein level was similar to RBPMS2. Thus, RBPMS overexpression was sufficient to promote SMC splicing patterns not only in the SM cell line, PAC1, but also in two unrelated human cell lines with similar effects.

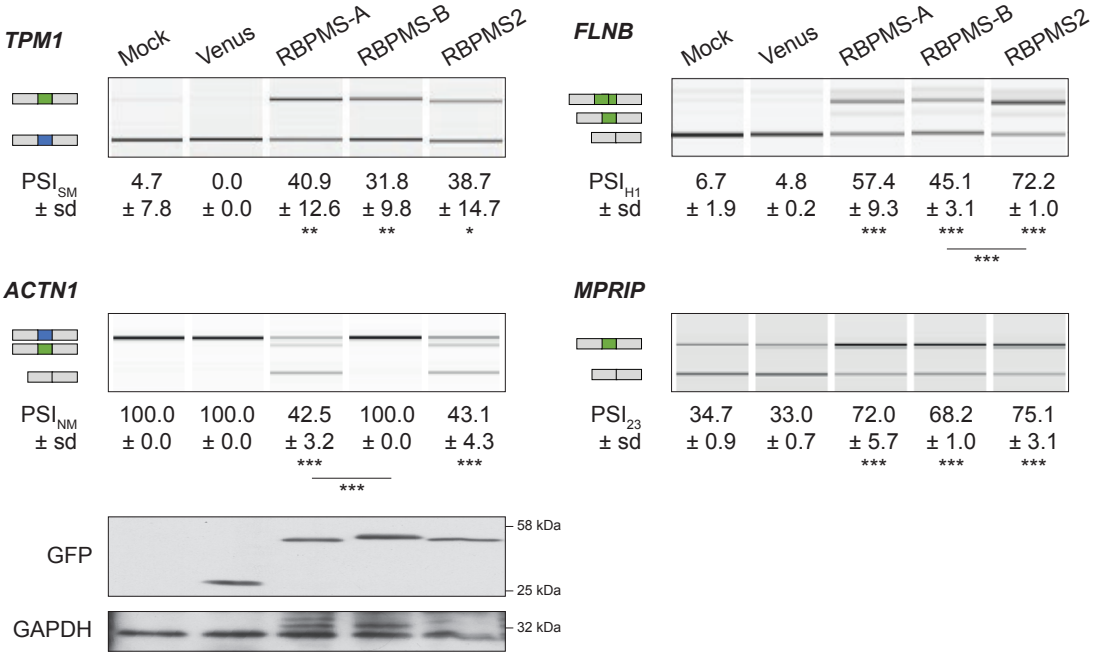


Fig. 3.13 RBPMS switches the SM differentiated splicing in HEK293 cells Venus tagged RBPMS isoforms and RBPMS2 were overexpressed in HEK293 cells. Splicing of endogenous *TPM1*, *ACTN1*, *FLNB* and *MPRIP* were assessed by RT-PCR. Mock and Venus controls were also tested. Schematics of the splicing isoforms are shown on the left and indicate differentiated exons (green) and proliferative exons(blue). Mean ± sd (n= 3) of PSI values for *TPM1* exon 2, *ACTN1* exon NM, *FLNB* exon H1 and *MPRIP* exon 23 are shown below the gel images. Western blot anti-GFP for detection of Venus tagged proteins. GAPDH was used as a loading control. Minus RT and no template controls were also carried out for every RT-PCR (data not shown). Protein size markers are also indicated on the right side of the western blot. Statistical significance from a two tailed unpaired Student's t-test is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

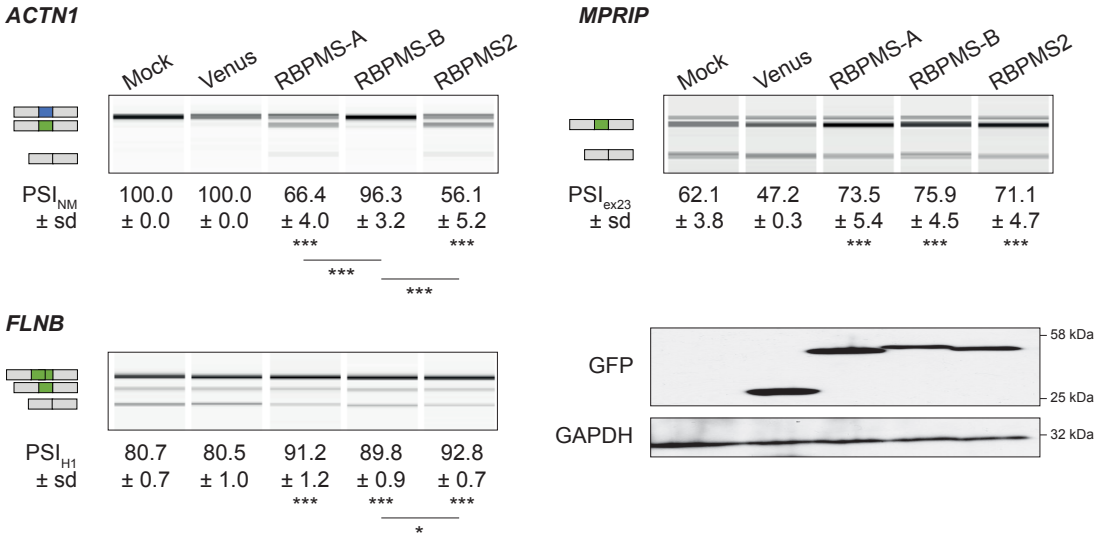


Fig. 3.14 RBPMS switches the SM differentiated splicing in HeLa cells Venus tagged RBPMS isoforms and RBPMS2 were overexpressed in HeLa cells. Splicing of endogenous *ACTN1*, *FLNB* and *MPRIP* were assessed by RT-PCR. Mock and Venus controls were also tested. Schematics of the splicing isoforms are shown on the left and indicate differentiated exons (green) and proliferative exons(blue). Mean ± sd (n= 3) of PSI values for *ACTN1* exon NM, *FLNB* exon H1 and *MPRIP* exon 23 are shown below the gel images. Western blot anti-GFP for detection of Venus tagged proteins. GAPDH was used as a loading control. Minus RT and no template controls were also carried out for every RT-PCR (data not shown). Protein size markers are indicated on the right side of the western blots. Statistical significance from a two tailed unpaired Student's t-test is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

3.3 Discussion

3.3.1 Identification of SMC master regulators via association of RBP genes with SM super-enhancers

By using the approach of identifying RBPs associated with super-enhancers in SMCs, nine candidate master regulators were found. The list includes a cleavage and polyadenylation factor (CPSF4L), a putative RNA helicase (DHX16), a U3 snoRNP protein (IMP3), a core U6 snRNP protein (LSM4), an RBP with an RNA Recognition motif - RRM (RBPMS), ribosomal proteins (RPL10A and RPL11), a histone modifier (SETD1B), and a scaffolding protein involved in RNA-mediated gene silencing (TNRC6C). Although RBPMS was the most promising candidate, there could be scope for critical roles played by the other RBPs. Nevertheless, it is hard to expect some essential role for the RBPs that were expressed at a very low level in adult SMC (Figure 3.3). Moreover, the candidates with substantial level of expression did not show significant changes between the differentiated and proliferative states nor were they more expressed in the latter.

3.3.2 RBPMS a potential master splicing regulator

RBPMS was the only RBP candidate whose expression level was regulated during SMC dedifferentiation, being more expressed in the differentiated state (Figure 3.3). Even though previous studies consistently confirmed RBPMS expression in the retina, heart, kidney, liver, and lungs by different methodologies (Farazi et al., 2014; Kwong et al., 2010; Shimamoto et al., 1996), human *Rbpms* seems to be expressed in a tissue-specific manner as indicated by RBPMS GTEX RNA-Seq data from multiple tissues and individuals (Figure 3.5). Additionally, RBPMS nuclear localization in the PAC1 cells was consistent with previous studies (Farazi et al., 2014; Fu and Ares, 2014; Kwong et al., 2010; Sagnol et al., 2014) and provided further support for its role as a splicing regulator (Figure 3.6). In that way, RBPMS was chosen as the main master regulator candidate to be tested. As seen in Figure 3.12, RBPMS regulated the splicing of SMC minigene reporters, promoting the differentiated state. Unlike other RBPs found to regulate these splicing markers (PTBP1 and MBNL1) (Gooding et al., 2013; Gromak et al., 2003), RBPMS was sufficient to cause the switch to the SMC isoform. More surprisingly, ectopic expression in non-smooth muscle cell lines could also recapitulate the SMC splicing (Figure 3.14 and 3.13). Once again, these are all features mentioned by Jangi and Sharp (2014) as characteristics of master splicing regulators. Additionally,

it is very likely that a tissue specific transcriptome requires an interplay between RBPs, so it is very interesting that RBPMS regulates splicing events like *Tpm1* and *Actn1* previously shown to be regulated by other RBPs, PTBP1 and MBNL1, highlighting a potential cooperative interaction between them.

RBPMS is an RBP that contains a single RRM responsible for protein dimerization and binding to CAC motifs (Farazi et al., 2014; Sagnol et al., 2014). As stated by its name (Shimamoto et al., 1996), RBPMS transcripts are highly regulated by AS, leading to multiple transcript isoforms, from which some are predicted to be NMD targeted. Therefore, conversely to what has been suggested for a master regulator (Jangi and Sharp, 2014), RBPMS could potentially be regulated by other splicing regulators and that yet remains to be determined. On the other hand, the two main isoforms, RBPMS A and B, did not seem to be differentially regulated in the PAC1 cells, as shown by the western blots against RBPMS in the differentiated and proliferative states (Figure 3.6).

In terms of functions, RBPMS has mainly been described in cytoplasmic roles. Its range of functions spans from formation of granules upon stress to transcriptional co-activator by regulating SMAD proteins localization (Farazi et al., 2014; Fu et al., 2015; Sun et al., 2006). Moreover, RBPMS2 has been studied as a dedifferentiation factor in gastrointestinal SMCs (Notarnicola et al., 2012; Sagnol et al., 2014), contrary to what is being proposed for RBPMS in here. In contrast, RBPMS2 was expressed at very low levels in VSMC, not showing upregulation upon dedifferentiation in rat aorta tissue (see Figure 3.4). Nevertheless, RBPMS2 ectopic expression resulted in regulation of splicing at a similar level to RBPMS, suggesting that both proteins share similar functions. A global study involving RBPMS overexpression in HEK293 cells revealed its RNA recognition element by PAR-CLIP but did not uncover any role in mRNA AS or stability after RNAseq analysis (Farazi et al., 2014). However, re-analysis of the RNAseq dataset by Dr. Miriam Llorian unraveled a few significant events regulated by RBPMS such as *ACTN1*, *FLNB* and *MPRIP*. Detection of events regulated by RBPMS overexpression in HEK293 cells in the reanalysis of the dataset is probably explained by two factors. First, a better RNAseq splicing analysis software was used allowing detection of more ASE. Secondly, in the original study events identified differentially spliced by RNAseq were discarded since they were not classified as direct targets according to the PAR-CLIP tags. Nonetheless, the events re-identified by rMATS expanded RBPMS activity range from *Tpm1* repressor to also an activator of splicing as seen in the regulation of *FLNB* exon H1 and *MPRIP* exon 23 (Figure 3.13 and 3.14). Interestingly, manual inspection of the intronic sequences flanking the AS exons, revealed conserved CACs, indicating a potential position-dependent activity. Finally,

RBPMS isoform B which was the least active in the splicing of *Tpm1* and *Actn1*, surprisingly showed to have similar activity to RBPMS-A and RBPMS2 when acting as an activator (Figures 3.14 and 3.13). These isoforms are very conserved in their N-termini, varying only in the very C-terminus (Figure 3.15). Therefore, it is expected that the isoform-specific amino acid sequence determines its activity. Nevertheless, it remains to be determined whether RBPMS-A residues confers it a repressor role or if RBPMS-B amino acid tail inhibits its repressive function.

Fig. 3.15 RBPMS-A and RBPMS-B vary in their C-termini. Multiple alignment of rat RBPMS isoforms A and B was generated using CLUSTAL O (1.2.4) and visualized with MView. Amino acids conserved with the reference sequence, RBPMS-A, were colored (bright green - hydrophobic, dark green - large hydrophobic, purple - polar, red - positive charge, dark blue - negative charge, bright blue - small alcohol, yellow - cysteine).

3.4 Final conclusions

This chapter aimed to identify master splicing regulators in SMCs via association of RBP genes with super-enhancers as well as initially validate the candidates by testing regulation upon some SM-specific AS events. Out of nine RBPs identified associated with SM super-enhancers, RBPMS was the most promising candidate, being predominantly nuclear and highly expressed in the differentiated SMC but strongly down-regulated in proliferative SMC. This feature of *Rbpms* met one of the criteria of a master splicing regulator described in Jangi and Sharp (2014): wide dynamic range of expression. Furthermore, RBPMS overexpression was sufficient to switch the splicing isoforms of *Tpm1* and *Actn1* towards the SM specific pattern in the PAC1 cells. More

surprisingly, RBPMS also promoted the SM differentiated patterns in unrelated cells, HeLa and HEK293 cells. In these cell lines, it was possible to screen for endogenous splicing changes, which revealed RBPMS role as an activator of *FlnB* exon H1 and *Mprip* exon 23. Finally, overexpression of the different isoforms, RBPMS A and B, shed some light onto differential activity between isoforms, with isoform B being the least active. Therefore, these data established a new function for RBPMS as a splicing regulator of a few SMC ASEs and supported further analysis for the identification of RBPMS targets in SMCs by global approaches such as RNAseq of PAC1 cells where RBPMS levels were manipulated either by knockdown and overexpression (see Chapter 4).

In conclusion, this chapter has the following points as the main findings:

1. Nine RBP genes were found associated with human SM super-enhancers: CPSF4L, DHX16, IMP3, LSM4, RBPMS, RPL10A, RPL11, SETD1B and TNRC6C.

2. *Rbpms* expression is regulated during rat aorta and PAC1 cells dedifferentiation, being more expressed in the differentiated state.

3. RBPMS predominantly localized to the nucleus but was also present in the cytoplasm.

4. RBPMS recognition elements, CAC motifs, were conserved upstream of SM proliferative exons such as *Tpm1* and *Actn1* and downstream of SM differentiated exons.

5. RBPMS promotes the SMC splicing pattern in the rat SM cell line PAC1 cells, switching the splicing isoforms of the *Tpm1* and *Actn1* minigene reporters towards the differentiated transcripts.

6. RBPMS promotes the SMC splicing pattern in non-SM cell lines, HeLa and HEK293, switching the splicing of endogenous transcripts, e.g. repressing the NM exon in *TPM1* and potentially also in *ACTN1*, and activating SM exons in *FLNB* and *MPRIP*.

7. *Rbpms* was found to be expressed in several isoforms in PAC1 cells, being RBPMS-A and B the major protein isoforms.

8. RBPMS A was the most active isoform together with its paralog, RBPMS2.

In addition to that, this chapter also led to other questions listed below:

- **What are the other mRNA targets of RBPMS?**
- What are the cytoplasmic functions of RBPMS?
- How are the different isoforms regulated?
- **Does RBPMS repress the exon NM or activate exon SM in *Actn1* splicing?**
- **Does RBPMS directly regulate splicing of these targets?**
- **What is the importance of the conserved CACs in the RBPMS regulation of the SMC splicing?**
- **What is the mechanism behind RBPMS regulation of splicing?**
- **How do the isoforms display different activities in splicing?**
- **What are functional consequences of the splicing changes regulated by RBPMS to the SMC phenotype?**

The questions highlighted in bold font were addressed in the following chapters.

Chapter 4

RBPMS: a regulator of the differentiated SMC splicing program

4.1 Introduction

4.1.1 RBPMS

The RNA-Binding Protein with Multiple Splicing (RBPMS) protein family was first described in a human library screening (Shimamoto et al., 1996) and it was also subsequently reported in a second screening using *X. laevis* embryonic heart development library (Gerber et al., 1999). Thereby, RBPMS is also known as HERMES which stands for HEart RRM Expressed Sequence. However, the Xenopus HERMES, originally thought to be RBPMS ortholog, is now known to be RBPMS2. As a result, in some species HERMES might refer to RBPMS2 instead. RBPMS protein family consists of RBPs containing a single RRM (RNA recognition motif), an RNA-binding domain presented in a large number of other RBPs (Ascano et al., 2012; Farazi et al., 2014). In humans, the family consists of RBPMS and its paralog, RBPMS2, which share 67% amino acid identity and variation observed mostly in the N and C-termini. RBPMS, as indicated by its name, is highly alternatively spliced, generating several mRNA isoforms that result in multiple protein isoforms (Shimamoto et al., 1996). For humans, four RefSeq RBPMS mRNA transcripts are annotated encoding three protein isoforms that are very similar in the majority of their sequence. On the other hand, RBPMS2 seems to be predominantly expressed as one single isoform. In rats, only one Refseq transcript isoform is annotated for RBPMS, however the rat genome is not

well annotated. For instance when searching for RBPMS in mouse, a close relative of rat with a better annotated genome, several isoforms similar to human isoforms were found. Indeed in this study two major RBPMS isoforms, A and B, were isolated from rat PAC1 cells. Human and rat RBPMS proteins are very similar displaying ∼99% amino acid identity. In fact, the RBPMS family is conserved across vertebrates (Figure 4.1) and orthologs, CPO and MEC-8, are also found in *Drosophila* and *C. elegans* respectively.

RBPMS was found associated with transcriptional super-enhancers in three different SM tissues in Chapter 3. These large clusters of transcriptional enhancers drive high transcription of genes important for cell identity (Hnisz et al., 2013). Although the proximity to this elements could explain RBPMS expression level in differentiated SMC, the constellation of factors that occupy these super-enhancers remains unknown. Investigation of RBPMS transcriptional control is limited to a single study in HEK293 cells (Ye et al., 2018), which revealed the importance of Sp1 binding sites (GC boxes) for RBPMS transcriptional regulation. However, these experiments were performed in a cell line in which RBPMS is not very abundant. Therefore, the role of these transcription factor binding sites is expected to vary in different cell types.

RRM containing RBPs are known to be involved in a range of RNA processes, for example pre-mRNA splicing, transport, localization and stability (Maris et al., 2005). Indeed, suggested functions of RBPMS involve mRNA stability, transport, localization in stress granules and also translation, as well as heart and kidney development, *X. laevis* oocyte maturation, retinal differentiation and lastly co-transcriptional activation (Farazi et al., 2014; Furukawa et al., 2015; Gerber et al., 2002; Hörnberg et al., 2016, 2013; Sun et al., 2006). RBPMS functions described in previous studies are summarized in Figure 4.2.

A systematic study of RBPMS in HEK293 cells defined its transcriptome-wide mRNA targets using photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) (Farazi et al., 2014). Further biochemical and computational approaches resolved the RBPMS RRM RNA-binding site as tandem CAC trinucleotide motifs separated by variable spacer length (1 to 12 nt) (Farazi et al., 2014). RBPMS mainly bound to the 3′ UTR and intron regions, but it also bound coding and repeat elements. Additionally, the structure of RBPMS (1.79 Å) and its paralog RBPMS2 (Sagnol et al., 2014; Teplova et al., 2016) uncovered the basis of the CAC RNA recognition as well as the dimeric nature of the RBPMS RRM. The dimerization state is then consistent with the PAR-CLIP detected RBPMS binding

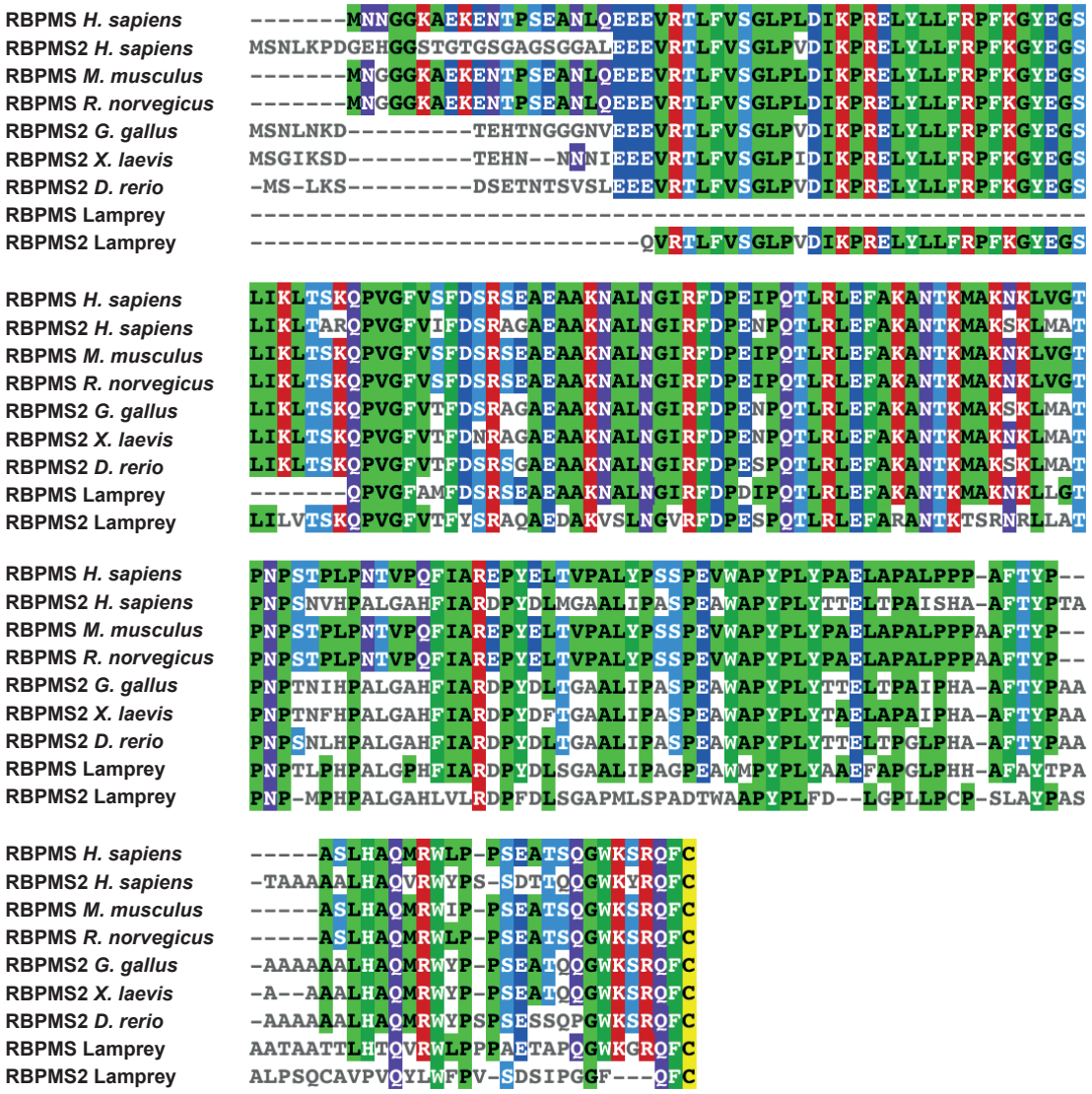


Fig. 4.1 RBPMS is conserved in vertebrates. Multiple alignment of RBPMS-A and RBPMS2 protein sequences across different vertebrate species was generated using CLUSTAL O (1.2.4) and visualized with MView. Only amino acids conserved with the reference sequence, RBPMS *H. sapiens*, were colored (bright green - hydrophobic, dark green - large hydrophobic, purple - polar, red - positive charge, dark blue - negative charge, bright blue - small alcohol, yellow - cysteine).

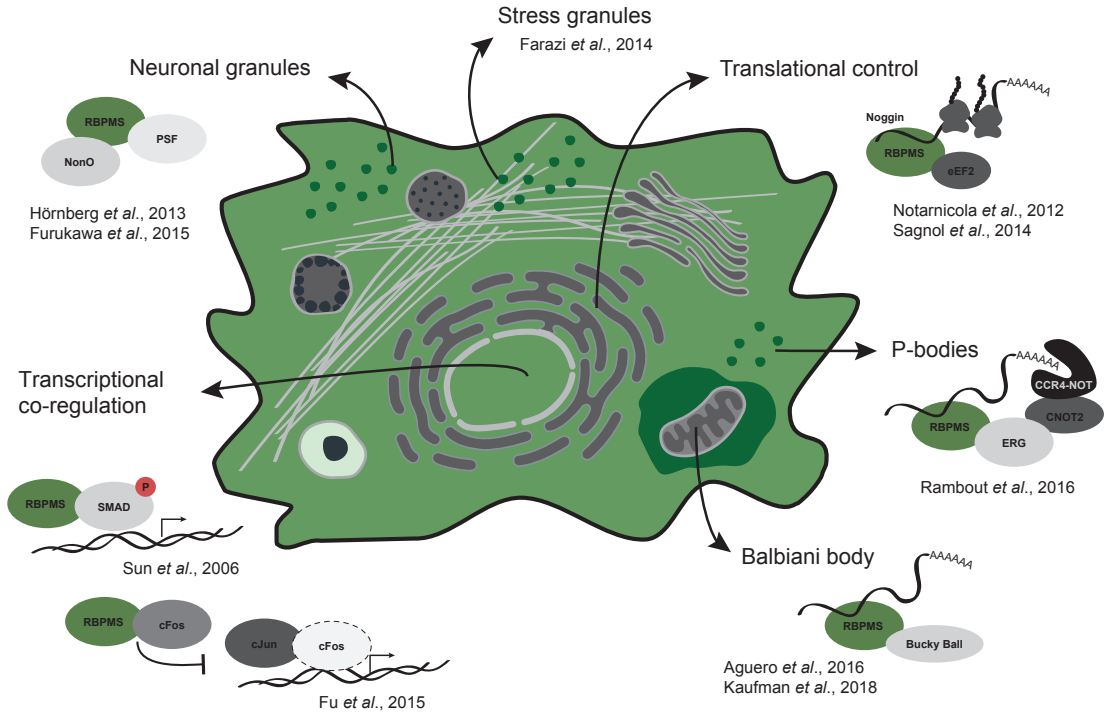


Fig. 4.2 RBPMS protein family functions. Summary of RBPMS and RBPMS2 reported to date. Green color indicate RBPMS localization obtained from COMPARTMENTS (Binder et al., 2014).

sites comprised of tandem CAC repeats. These structural features are further discussed in following chapters.

Therefore, despite all the studies involving RBPMS and RBPMS2 proteins in different cell types and stages of development, the nuclear role of RBPMS in adult VSMCs was still neglected. In this chapter, motivated by the initial results observed in Chapter 3, we sought to uncover RBPMS potential function in the regulation of mRNA abundance and more specifically AS in the VSMC phenotypic switching by manipulating its levels in the differentiated and proliferative PAC1 cell states followed by mRNAseq.

4.2 Results

4.2.1 Manipulation of RBPMS levels

To investigate other targets of RBPMS post-transcriptional control in a more global manner, RBPMS levels were manipulated by siRNA knockdown in differentiated PAC1 cells and lentiviral inducible RBPMS-A overexpression stable cell lines were produced using proliferative PAC1 cells (Figure 4.3). The latter was obtained by Clare Gooding in the laboratory.

RBPMS protein levels were verified by western blot after siRNA treatment and induction of overexpression with doxycycline (Figure 4.3). RBPMS knockdown resulted in 75% reduction of the protein levels of both isoforms with no observed changes in the SMC marker, ACTA2. Moreover, induction of the overexpression of the 3xFLAG tagged RBPMS-A showed to be specific to the RBPMSA cell lines and to the doxycycline treatment. In agreement with the pattern of RBPMS expression in PAC1 cells, no endogenous RBPMS was detected in the proliferative lentiviral cell lines.

Next, changes in endogenous *Flnb* and *Actn1* splicing due to RBPMS levels manipulation were assessed by RT-PCR (Figure 4.4). For *Flnb*, RBPMS knockdown decreased the level of exon H1 inclusion to nearly half of the observed in the control treatment (94.8% to 52.3%) whereas overexpression nearly promoted its complete inclusion (PSI= 93.3%). For the *Actn1* mutually exclusive pair of exons, a similar pattern was observed, where RBPMS knockdown reduced the smooth muscle exon inclusion and RBPMS overexpression increased the SM exon inclusion (23.4% and 84.0%). In both splicing examples, RBPMS down and upregulation resulted in significant reciprocal changes, switching the splicing isoforms towards the opposite cell state pattern. Furthermore, the splicing levels from the knockdown and overexpression controls, D Ctr and P Ctr,

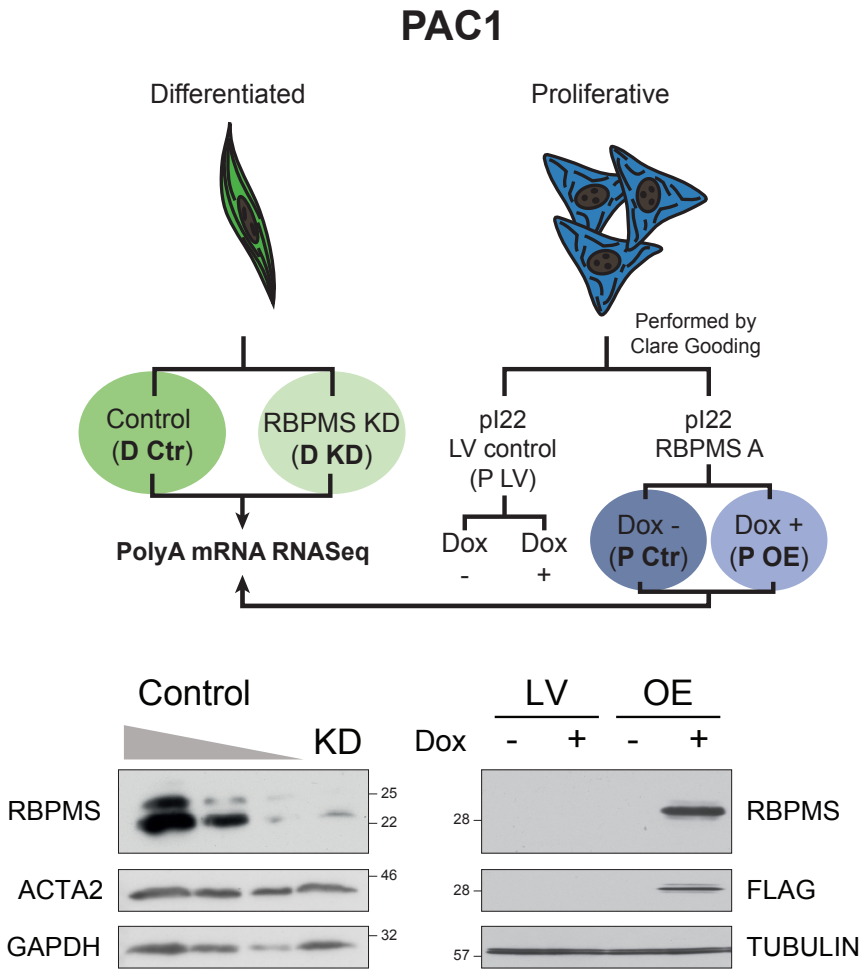


Fig. 4.3 Manipulation of RBPMS levels by knockdown and overexpression. **Top**, schematic of the experimental designs of RBPMS knockdown and overexpression in PAC1 cells. **Bottom**, western blots anti-RBPMS and anti-FLAG for detection of endogenous and overexpressed RBPMS respectively. ACTA2 was used as a SMC differentiation marker and GAPDH and TUBULIN were used as loading controls. Protein size markers are also indicated on the side of the western blots.

supported the different PAC1 cell states used in each experiment, differentiated and proliferative respectively.

In parallel to the RT-PCR analysis of the RBPMS overexpression regulated ASE, Clare Gooding in the laboratory also validated the lentiviral control cell lines (P LV) that contain the empty pInd22 construct. No splicing changes were observed upon doxycycline induction (Appendix - Figure A.1). Thus, these data ensured that the alterations in splicing observed in the P OE were due to the overexpression of RBPMS-A.

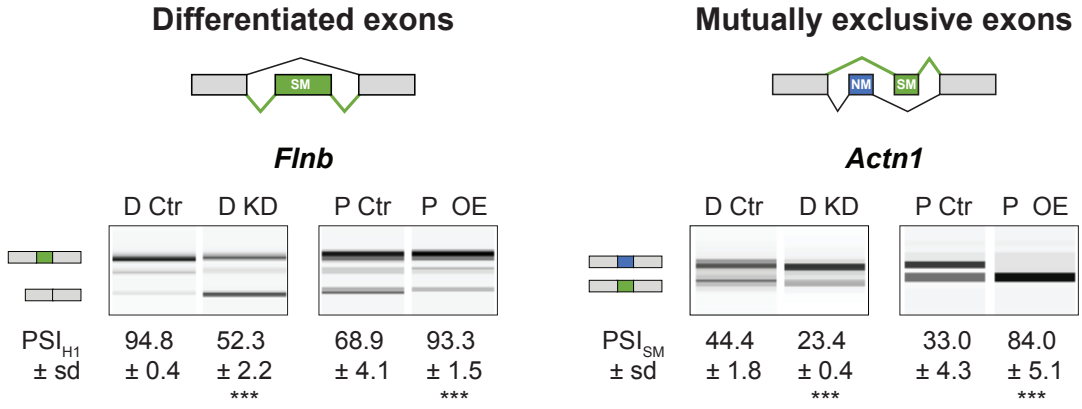


Fig. 4.4 RBPMS knockdown and overexpression affected *Flnb* and *Actn1* splicing. RT-PCR validation of endogenous *Flnb* exon H1 and *Actn1* mutually exclusive exons upon RBPMS knockdown and overexpression. Schematics of the splicing isoforms are shown on the left and indicate differentiated exons (green) and proliferative exon (blue). PSI values are shown as mean ± sd (n= 3). A minus RT and a no template control were also carried out for every RT-PCR (data not shown). Statistical significance from a two tailed Student's t-test is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

To further verify whether the effects of RBPMS knockdown were reproducible, a second siRNA was tested alongside (Figure 4.5). Both siRNAs efficiently reduced RBPMS protein levels to a similar degree, below 75% compared to cells treated with control siRNA (Figure 4.5). Moreover, similar to the previous knockdown results, RBPMS silencing by both siRNAs led to changes in *Flnb* and *Actn1* splicing, as well as *Tpm1* (Figure 4.5). In all the ASE assessed, both knockdowns resulted in equivalent effects and again promoted the non smooth muscle isoforms.

The role of RBPMS2 in the regulation of PAC1 cells AS and its possible compensation upon RBPMS silencing were investigated by RBPMS2 siRNA knockdown

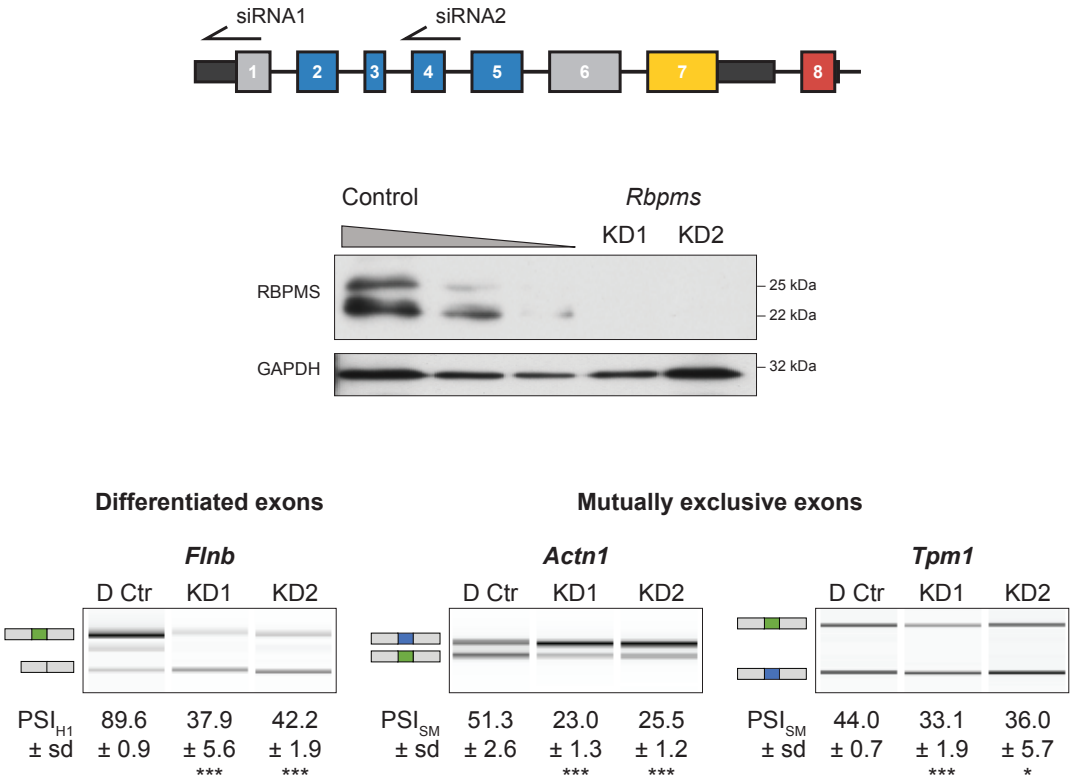


Fig. 4.5 RBPMS knockdown with a second siRNA. A Western blot anti-RBPMS for detection of endogenous RBPMS upon the knockdown with two different siRNAs (KD1 and KD2). Control sample in a serial dilution (1, 1:2, 1:4). GAPDH was used as a loading control. Protein size markers are also indicated on the side of the western blots. **B** RT-PCR validation of differentiated and proliferative exons regulated upon RBPMS knockdown. Schematics of the splicing isoforms are shown on the left and indicate differentiated exons (green) and proliferative exon (blue). PSI values for the SM exons are shown as mean ± sd (n= 3). A minus RT and a no template control were also carried out for every RT-PCR (data not shown). Statistical significance from a two tailed unpaired Student's t-test is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

(Figure 4.6). In Chapter 3, it was shown that RBPMS2 mRNA expression was not affected during PAC1 dedifferentiation and its overexpression revealed a potential role in splicing regulation. However, in aorta tissue, RBPMS2 transcript abundance was much lower than RBPMS. Thus, in agreement with that, despite an ∼80% reduction of RBPMS2 mRNAs (Figure 4.6A), minimal effects to splicing of *Actn1* and *Tpm1* were observed (Figure 4.6B). Additionally, no cross-regulation between paralogs was revealed by the qRT-PCR. Lastly, manipulation of both RBPMS and RBPMS2 levels did not show further effects to the ASE than the observed with RBPMS only knockdown. Therefore, this eliminated the need for a double knockdown in order to achieve maximal effects of RBPMS in splicing.

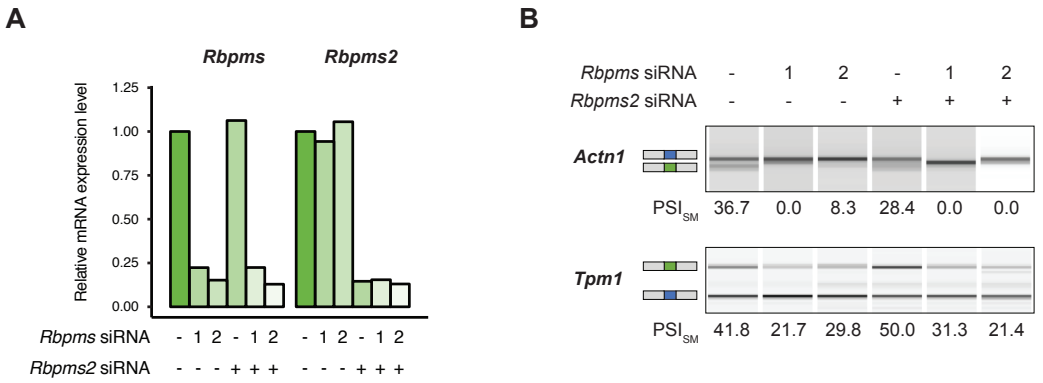


Fig. 4.6 RBPMS2 knockdown in PAC1 cells has minimal effects on *Actn1* and *Tpm1* splicing. Top qRT-PCR assessing transcripts level after RBPMS, RBPMS2 and combined knockdown. GAPDH was used a housekeeper gene for normalization. Number 1 and 2 indicate RBPMS siRNA1 and siRNA2 respectively. **Bottom** RT-PCR validation of endogenous *Actn1* and *Tpm1* mutually exclusive exons upon knockdown. Schematics of the splicing isoforms are shown on the left and indicate differentiated exons (green) and proliferative exon (blue). Experiment of n= 1, so no sd or statistical test shown.

Having confirmed the efficacy of RBPMS knockdown and overexpression, RNA samples were prepared for polyA mRNA sequencing. Samples were from three RBPMS knockdowns (siRNA1) in differentiated PAC1 cells and three independent RBPMS-A inducible lentiviral populations of proliferative PAC1, corresponding to the experiments shown in Figures 4.3 and 4.4. Moreover, comparison of the knockdown and overexpression controls provided insights on the PAC1 transcriptional and AS program during dedifferentiation.

4.2.2 RNAseq of RBPMS knockdown and overexpression

cDNA libraries and mRNA sequencing of RBPMS knockdown and RBPMS-A overexpression in PAC1 cells were carried out by the CRUK Genomics Facility. Prior to sequencing, RNA and cDNA libraries were controlled for their qualities. RNAseq data were then primarily analyzed by Dr Miriam Llorian who also performed the mRNA abundance and splicing analyses. Further data exploration and visualization were carried out by me using R Studio.

4.2.2.1 mRNA abundance analysis

Analysis of differential mRNA abundance was carried out by DESeq2 (Love et al., 2014) and changes were analyzed for PAC1 dedifferentiation (D Ctr vs P Ctr), RBPMS knockdown (D KD vs D Ctr) and RBPMS-A overexpression (P OE vs P Ctr), Figure 4.7. First, as expected, RBPMS knockdown and overexpression reduced and increased RBPMS levels by 0.6 and 70 folds respectively. In addition to that, RBPMS transcript abundance was ∼4 fold more abundant in the differentiated than in the proliferative state. In terms of global transcripts regulated, the differentiated and proliferative PAC1 cell comparison showed the greatest number of genes differently abundant (2274 genes), followed by overexpression (725) and knockdown (192). Furthermore, a bias towards reduction of mRNA abundance was observed in the D Ctr vs P Ctr (1444 down and 830 upregulated) whereas the RBPMS experiments, mainly overexpression, showed genes more associated with an increased level of transcripts. ∼78% of the genes differentially abundant in RBPMS-A overexpression were upregulated. Moreover, although high levels of *Rbpms* and *Acta2* co-existed in the differentiated state of PAC1 cells, no changes in *Acta2* or other SMC markers, such as *Smtn* and *Cnn1* (data not shown), were detected upon manipulation of RBPMS expression (Figure 4.7).

The top ten up and downregulated genes during PAC1 dedifferentiation, RBPMS knockdown and overexpression are found in Table 4.1. Apart from LOC100911692 in RBPMS overexpression and knockdown, little overlap was observed in the top expressed genes between all comparisons. In addition to that, motivated by the similar profile of up regulated genes between the PAC1 dedifferentiation and overexpression, correlation was verified for the genes showing differential abundance in the different experiments (Figure 4.8). As previously highlighted in the table for top regulated genes (Table 4.1), little overlap was observed between genes with changes in mRNA abundance in the different experiments. RBPMS knockdown showed the best correlation, -0.64, but no significant relationship was observed between RBPMS experiments. Thus, changes in

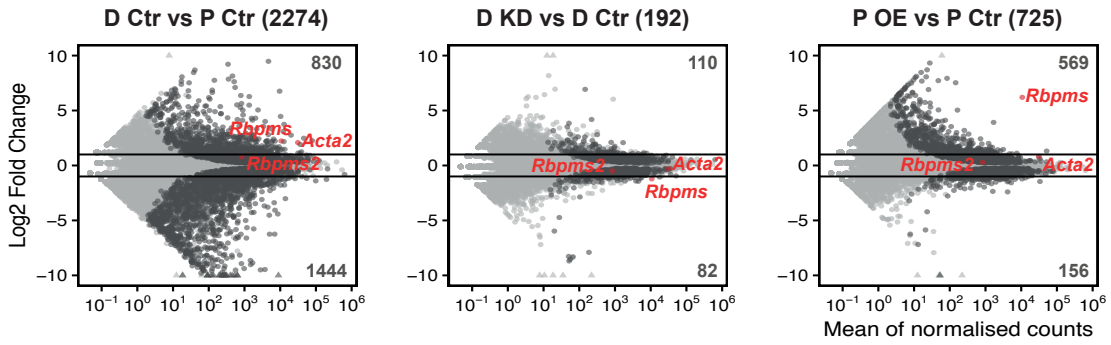


Fig. 4.7 RBPMS is associated with a few changes in mRNA abundance MA plots of the changes in mRNA abundance in PAC1 dedifferentiation (D Ctr vs P Ctr), RBPMS knockdown (D KD vs D Ctr) and RBPMS overexpression (P OE vs P Ctr) detected by DESeq2. Genes with significant changes (p-adj<0.05) in mRNA abundance are highlighted in dark gray and genes with non-significant changes (p-adj≥0.05) are shown in light gray. *Rbpms*, *Rbpms2* and *Acta2*, a SMC marker, genes are indicated in red. Numbers at the top and bottom correspond to the number of genes upregulated and downregulated, respectively. Horizontal lines in the MA plots represent log2 fold change = 1 and -1.

mRNA abundance associated with the manipulation of RBPMS levels seem to be a minor effect, due to its variability between knockdown and overexpression. Moreover, it is hard to draw any biological importance of these genes as little overlap was observed with the PAC1 dedifferentiation regulated genes.

To address the consequences of the mRNA abundance changes to the samples grouping, a Principal Component Analysis (PCA) based on mRNA abundance was performed (Figure 4.9). Differentiated and proliferative samples were clearly separated by PC1, 82% variance. The differences between the independent lentiviral transductions for the generation of the overexpression cell lines are reflected by PC2, 10% variance. Consistent with the previous analyses, the PCA results support the fact that manipulation of RBPMS levels is not sufficient to overcome the differences between differentiated and proliferative PAC1 cells at the level of mRNA abundance.

4.2.2.2 AS analysis

Analysis of mRNA alternative splicing was carried out by rMATS (Shen et al., 2014) and changes were analyzed for PAC1 dedifferentiation (D Ctr - P Ctr), RBPMS knockdown (D KD - D Ctr) and RBPMS-A overexpression (P OE - P Ctr), Figure 4.10. ASEs were classified into five categories: skipped exons (SE), mutually exclusive

Table 4.1 Top ten upregulated and downregulated genes detected by DESeq2

PAC1 dedifferentiation		**RBPMS knockdown**		**RBPMS overexpression**	
Gene name	log2FC	Gene name	log2FC	Gene name	log2FC
Trpc6	9.5	*Cpne1*	6.9	*LOC100911692*	9.3
Galnt18	9.2	*Aldh3a1*	4.2	*Hells*	8.7
Clec4m	8.7	*LOC100909857*	3.8	*Cyp2d2*	8.5
AABR07035650.1	8.5	*Sprr1a*	3.5	*Chrne*	8.1
Wfdc21	8.1	*Me3*	3.2	*Ccdc87*	8.1
Hoxc6	8.1	*Slurp1*	2.8	*Cacna1d*	8.0
Spon1	7.9	*Kcnk2*	2.8	*Slc5a5*	7.9
Nckap1l	7.7	*Rtp4*	2.6	*Nyx*	7.8
Lin7a	7.7	*Il18rap*	2.6	*Tnfsf8*	7.7
MGC114427	7.6	*Dhrs9*	2.4	*Tekt5*	7.7
Ntrk1	-10.5	*Zfp804a*	-2.4	*Ttyh3*	-2.4
Rbm46	-10.5	*Hist2h2be*	-2.7	*Lima1*	-2.4
Tex19.2	-10.6	*Ak4*	-2.9	*Zfp93*	-2.6
Slc7a3	-10.9	*LOC100911692*	-4.7	*Lgr5*	-2.8
Mael	-11.2	*AABR07035375.1*	-6	*Mettl21b*	-3.2
Slc6a17	-11.3	*Tnks2*	-7.9	*LOC108348161*	-4.2
Aass	-11.4	*LOC102549115*	-8.3	*ACEA_U3*	-5.9
Adad2	-11.9	*Sbk1*	-8.3	*Cracr2b*	-6.2
Cdh5	-12.2	*LOC100909505*	-8.5	*Brms1*	-7.2
Tmem200b	-20.3	*Rbm12*	-8.7	*Rfxapl1*	-22.6

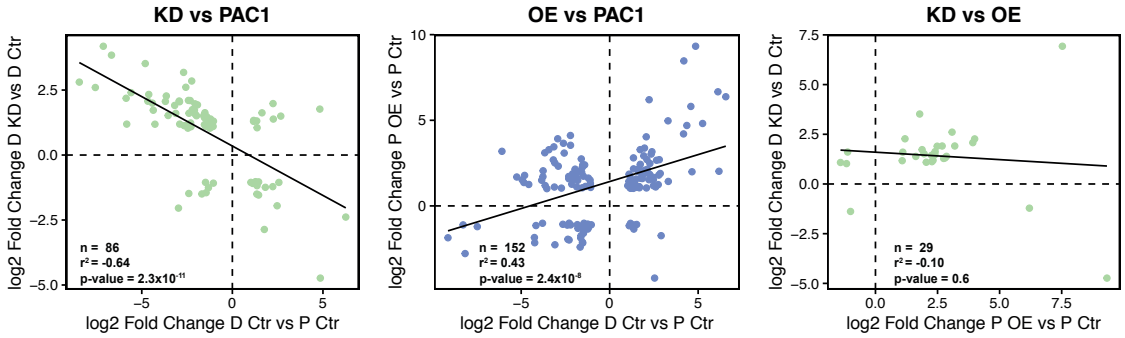


Fig. 4.8 Genes showing changes in mRNA abundance in the different comparisons display little overlap Fold change correlation of the overlapping genes across the different experiments: PAC1 dedifferentiation (PAC1), RBPMS knockdown (KD) and RBPMS-A overexpression (OE). Black lines indicate the linear regression model. Pearson correlation test results are shown in the plot. The correlation statistical significance and coefficient are indicated by the p-value and r^2 in the plot.

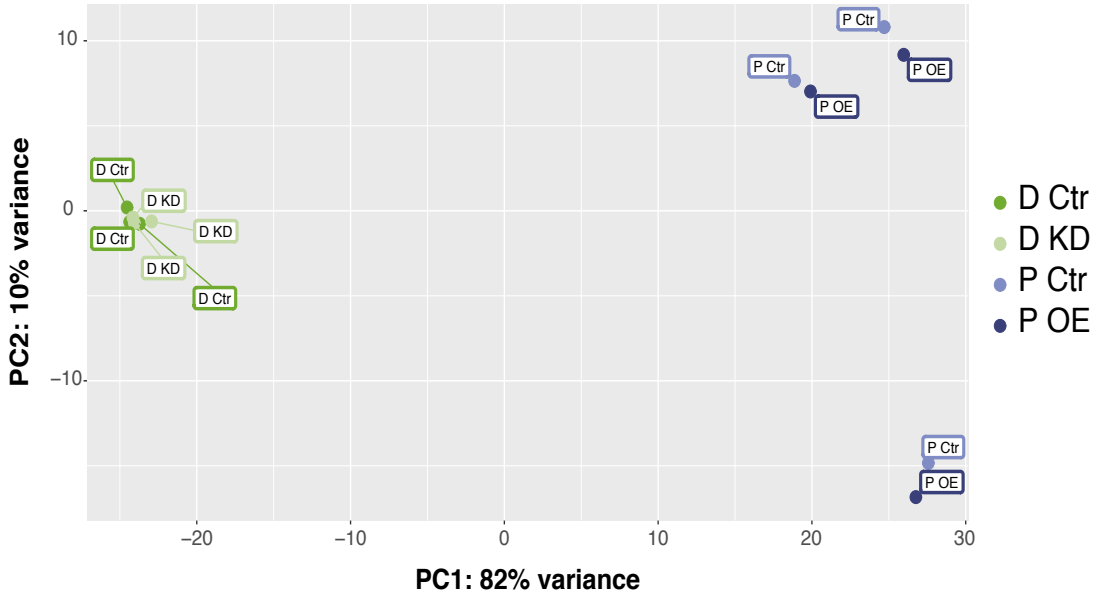


Fig. 4.9 Changes in mRNA abundance due to RBPMS manipulation do not overcome changes during PAC1 dedifferentiation. Principal Component Analysis (PCA) based on mRNA abundance variance of RBPMS knockdown (D Ctr and D KD) and overexpression (P Ctr and P OE).

exons (MXE), alternative 5′ and 3′ splice sites (A5SS and A3SS) and retained introns (RI) (Figure 4.10). ASE were considered differentially spliced if showing an FDR less than 0.05 and |ΔPSI| greater than 10%. First, the proportion of each type of ASE identified in the different experiments were verified (Figure 4.10). RBPMS knockdown and overexpression did not show any preference towards regulation of a specific type of AS, since its AS distribution was similar to the PAC1 dedifferentiation. Moreover, SE and MXE together corresponded to the majority of the regulated events (∼60%). Absolute numbers of the AS analysis uncovered the substantial effects of RBPMS as an activator and repressor of splicing in PAC1 cells (Figure 4.11). RBPMS knockdown regulated 318 ASEs, 136 more included and 182 more excluded exons, whereas RBPMS overexpression showed a much larger number in AS changes, 4934 ASEs of which 1286 more included and 3648 more excluded exons. During PAC1 cells dedifferentiation, 635 exons were regulated, 352 activated and 283 repressed exons. Given that RBPMS knockdown regulated only 2-fold less ASEs than PAC1 dedifferentiation, by far RBPMS manipulation affected more splicing than mRNA abundance in the PAC1 cells. On the other hand, RBPMS-A overexpression resulted in a strong regulation of AS, beyond 5-fold more ASE than during dedifferentiation of PAC1 cells. It is also notable that RBPMS acts more as a repressor as suggested by the skewed number of SE towards less inclusion. The top ten activated and repressed exons identified by rMATS during the PAC1 dedifferentiation, RBPMS knockdown and overexpression are found in Table 4.2, 4.3 and 4.4. Consistent with previous validation of RBPMS role in splicing (see Figure 4.4), *Actn1* and *Flnb* ASEs were among the top ten RBPMS regulated exons.

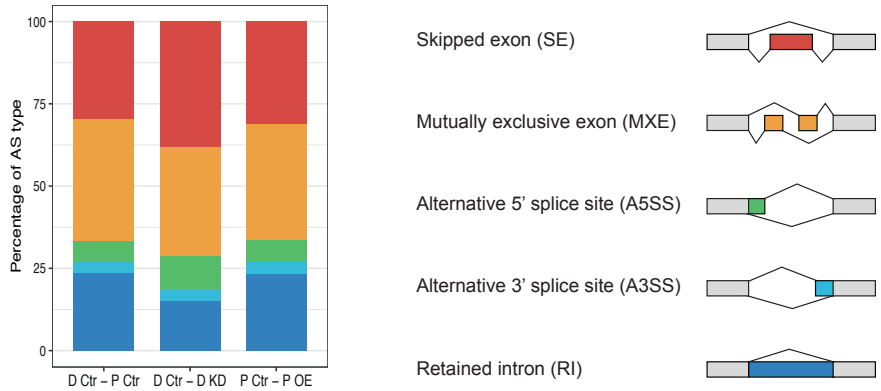


Fig. 4.10 RBPMS affects all types of AS with no preference to a specific type. Proportion of significant ASEs identified by rMATS (FDR < 0.05 and |ΔPSI| > 10%) for each type of splicing. Schematics of the five different AS types identified by rMATS are shown on the right.

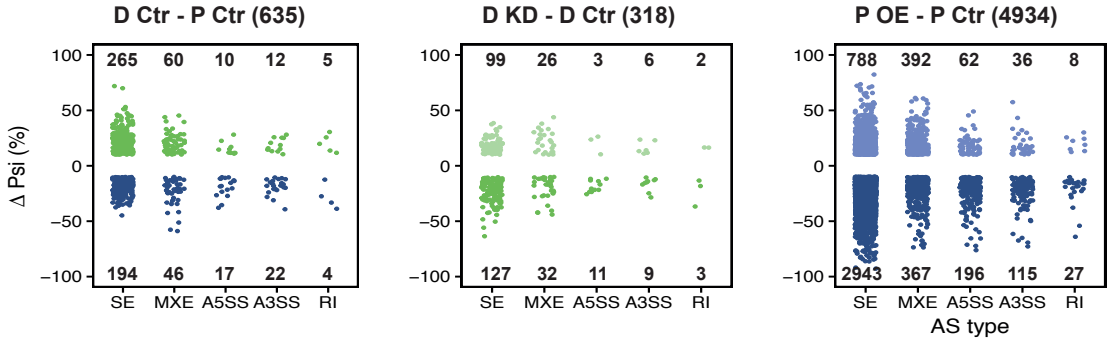


Fig. 4.11 RBPMS regulates AS in PAC1 cells. AS changes in PAC1 dedifferentiation (D Ctr - P Ctr), RBPMS knockdown (D KD - D Ctr) and RBPMS overexpression (P OE - P Ctr). Plots include only ASE that showed FDR < 0.05 and $|\Delta PSI|$ greater than 10%. ASEs were divided into Skipped Exon (SE), Mutually Exclusive Exon (MXE), Alternative 5′ and 3′ Splice Site (A5SS and A3SS) and Reteined Intron (RI) by rMATS. In each comparison, the colors indicate exons more included in the respective sample (see Figure 4.3). Numbers highlight the numbers of significant ASE found activated or repressed (top and bottom) in each type of event.

Table 4.2 Top ten activated and repressed exons in PAC1 dedifferentiation detected by rMATS

Gene name	ΔPSI	ASEID	AStype
Elovl1	71.9	5.137257786.137257979.137257479.137257708.137258385.137258445	SE
Elovl1	70.0	5.137257776.137257979.137257479.137257708.137258385.137258445	SE
Col4a5	53.0	X.112958131.112958140.112956286.112956472.112964677.112964811	SE
Osbpl6	51.2	3.63412201.63412294.63411343.63411447.63421091.63421257	SE
Fnip2	51.1	2.178273858.178274004.178272369.178272441.178274159.178274260	SE
Fmnl3	47.4	7.140993715.140993748.140992495.140992604.140996289.140996373	SE
Flnb	46.1	15.18780302.18780341.18779402.18779505.18783400.18783648	SE
Pdlim7	45.8	17.9659208.9659229.9658265.9658294.9659478.9659615	SE
Alg14	45.3	2.224915567.224915697.224916092.224916217.224886636.224886768.224930476.224931459	MXE
Mprip	45.2	10.46129845.46129896.46129339.46129425.46132876.46133571	SE
Axl	-37.5	1.82566746.82566773.82563906.82564039.82568169.82568380	SE
Fam193b	-37.9	17.9617725.9617866.9617725.9617770.9618366.9618606	A5SS
Zfyve19	-38.8	3.111015802.111016048.111015802.111015853.111015924.111016048	RI
Serpinh1	-39.2	1.164307482.164307705.164307482.164307640.164308246.164308317	A3SS
Pls3	-41.8	X.119065134.119065209.119076730.119076854.119030418.119030494.119078761.119078842	MXE
Actn1	-42.6	6.103379313.103379379.103379819.103379900.103378007.103378166.103380858.103381005	MXE
Zfp532	-44.8	18.61335535.61335698.61328527.61328704.61361848.61361961	SE
Nutf2	-51.2	19.37831236.37831400.37839121.37839238.37830916.37830949.37845351.37845478	MXE
Elovl1	-57.7	5.137257479.137257708.137257830.137257979.137255923.137256002.137258385.137258445	MXE
Elovl1	-58.9	5.137257479.137257708.137257786.137257979.137255923.137256002.137258385.137258445	MXE

Table 4.3 Top ten activated and repressed exons in RBPMS knockdown detected by rMATS

Gene name	ΔPSI	ASEID	AStype
Prnp	43.8	3.124519940.124520061.124527946.124528058.124518258.124518356.124529319.124531316	MXE
Pusl1	38.5	5.173336675.173336884.173336033.173336421.173337515.173337691	SE
Prnp	38.0	3.124519940.124520061.124527962.124528054.124518258.124518356.124529319.124531316	MXE
Mbnl1	37.4	2.150864085.150864121.150858305.150858459.150864801.150864896	SE
Slfn13	34.8	10.70337408.70337518.70336023.70337090.70342232.70342411	SE
Actn1	34.4	6.103379313.103379379.103379819.103379900.103378007.103378166.103380858.103381005	MXE
Tmem109	33.2	1.226929746.226929869.226932378.226932595.226928607.226928852.226935626.226935689	MXE
Mbnl1	32.9	2.150864085.150864121.150858305.150858459.150867718.150867793	SE
Arhgap17	32.0	1.193243546.193243595.193245127.193245237.193239039.193239872.193246163.193246715	MXE
Camk2g	31.3	15.3983819.3983888.3974330.3974374.3985004.3985049	SE
Hspg2	-42.1	5.155853492.155853543.155851503.155851664.155853919.155854048	SE
Itga7	-42.2	7.3363146.3363278.3363529.3363649.3362381.3362637.3364518.3364726	MXE
Nutf2	-42.3	19.37839121.37839238.37840879.37841075.37830916.37830949.37845351.37845478	MXE
Mprip	-42.5	10.46129845.46129896.46129339.46129425.46132876.46133571	SE
Itga7	-44.0	7.3363146.3363278.3363529.3363649.3362381.3362637.3364823.3365017	MXE
Dhdh	-44.3	1.101474797.101474961.101471873.101472126.101476777.101476867	SE
Flnb	-48.3	15.18780269.18780341.18779402.18779505.18783400.18783648	SE
Dhdh	-50.4	1.101474797.101474959.101471873.101472126.101476777.101476867	SE
Itga7	-55.9	7.3363146.3363278.3362564.3362637.3363529.3363649	SE
Flnb	-63.6	15.18780302.18780341.18779402.18779505.18783400.18783648	SE

Table 4.4 Top ten activated and repressed exons in RBPMS-A overexpression detected by rMATS

Gene name	ΔPSI	ASEID	AStype
Gkap1	82.3	17.6816561.6816708.6807553.6807576.6816946.6817048	SE
Actn1	73.6	6.103379313.103379379.103378007.103378166.103379819.103379900	SE
Prtfdc1	72.2	17.87929406.87929517.87928989.87929066.87931271.87931318	SE
Svil	71.6	17.55289394.55290249.55288179.55288214.55297975.55298582	SE
Mical3	70.6	4.153559892.153560000.153559120.153559307.153571656.153571830	SE
Prtfdc1	69.7	17.87929429.87929517.87928989.87929066.87931271.87931318	SE
G2e3	68.0	6.72099285.72099362.72097747.72097788.72099699.72099797	SE
Brd1	65.7	7.129690628.129690688.129690244.129690349.129696841.129697102	SE
Dennd1b	65.6	13.56120365.56120425.56115264.56115341.56121584.56121638	SE
Ptprf	65.2	5.137046364.137046397.137045483.137045581.137048659.137048775	SE
E2f1	-82.2	3.150064424.150064575.150062894.150064333.150064738.150064853	SE
Qtrt1	-82.8	8.22453468.22453550.22453278.22453388.22453991.22454439	SE
E2f1	-82.9	3.150064424.150064585.150062894.150064333.150064738.150064853	SE
Phldb2	-83.9	11.57457745.57457820.57454353.57454443.57462008.57462108	SE
Rnf44	-84.2	17.10468564.10468738.10467852.10468011.10469413.10469501	SE
Duoxa1	-84.6	3.114242245.114242455.114241056.114241591.114242701.114242915	SE
Brd4	-86.1	7.14230130.14230322.14228763.14229354.14230480.14230533	SE
Uckl1	-86.3	3.177066994.177067073.177066785.177066908.177067188.177067435	SE
Homer3	-91.9	16.20882892.20883024.20881517.20881783.20883116.20883227	SE
Homer3	-92.8	16.20882892.20883018.20881517.20881783.20883116.20883227	SE

To confirm RBPMS regulation of the ASEs identified in the RNAseq analysis, a subset of ASE was then validated by RT-PCR, using the same RNA samples originally used for the sequencing (Figure 4.12). Among the novel events validated, there were differentiated and proliferative cassette exons, some whose AS had previously been shown to be regulated in SMC, such as *Smtn* (Llorian et al., 2016). All the events tested by RT-PCR showed significant changes upon manipulation of RBPMS levels (Figure 4.12A). Although only the knockdown events are shown herein, C. Gooding, in the laboratory, validated the same events in the overexpression cell lines. In total, a subset of 28 ASEs was validated and all of them were found in good agreement with the RNAseq predicted ΔPSI (Figure 4.12B).

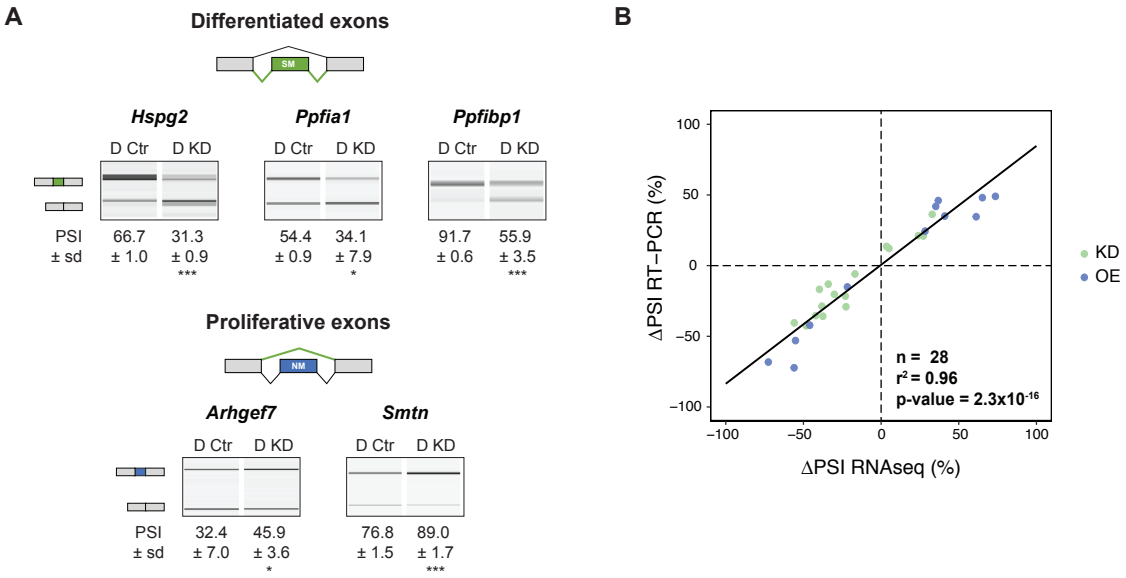


Fig. 4.12 RBPMS ASE ΔPSI values from RNAseq and RT-PCR highly correlated. A Validation of differentiated and proliferative cassette exons by RT-PCR. A minus RT and a no template control were also carried out for every RT-PCR (data not shown). Values shown as mean of the PSI ± sd (n = 3). Statistical significance was calculated using Student's t-test (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). **B** Correlation between ΔPSI values from rMATS prediction and RT-PCR validated experiments from RBPMS knockdown and overexpression (n=28). Black line indicates the linear regression model. Results from the Pearson correlation test are shown in the plot. The correlation statistical significance and coefficient are indicated by the p-value and r^2 in the plot.

Ptprf, *Piezo1* and *Itga7* are shown as other examples of differentiated and proliferative cassette exons as well as mutually exclusive exons (Figure 4.13). RT-PCR validations are shown along side Sashimi plots, which allow visualization of the RNAseq

data (Figure 4.13). Differentiated and proliferative exons were defined based on the inclusion levels in the two states in PAC1 cells. It is evident that both RBPMS knockdown and overexpression affected splicing consistently promoting the opposite effects. Furthermore, in all the three cases, RBPMS knockdown led to proliferative isoforms whereas overexpression of RBPMS-A regulated towards differentiated patterns.

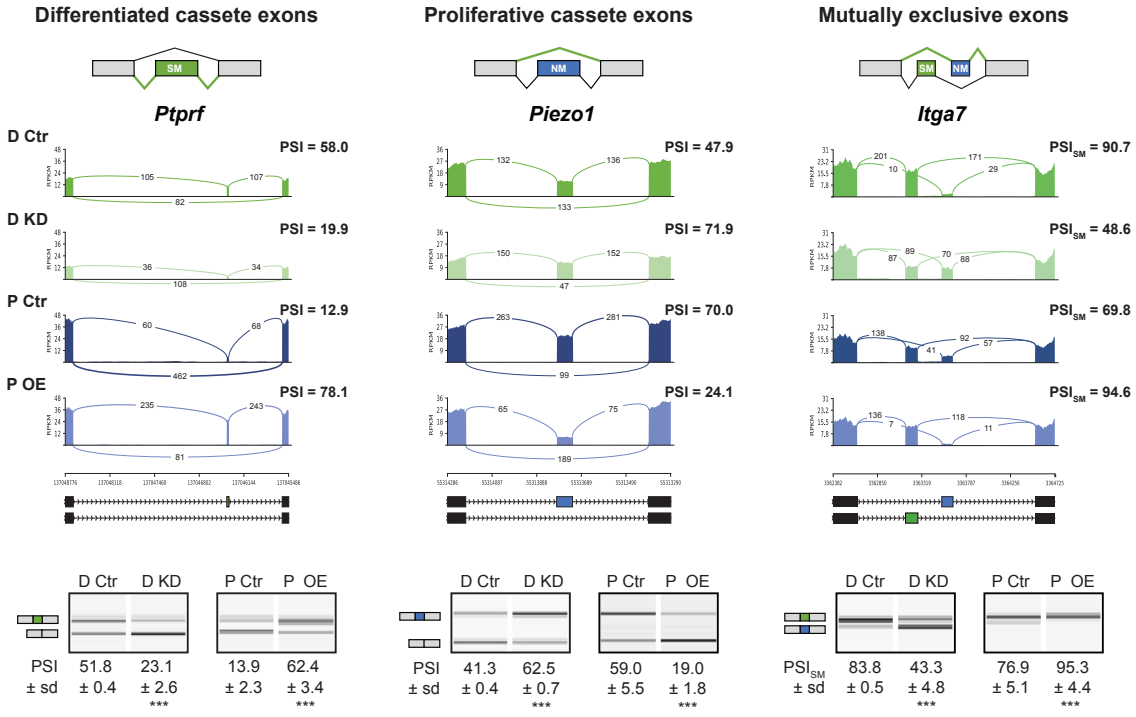


Fig. 4.13 RBPMS regulates splicing towards the differentiated state. Differentiated and proliferative cassette exons (green and blue, respectively) and mutually exclusive exons identified by rMATS were validated by RT-PCR. **Top**, sashimi plots for *Ptprf*, *Piezo* and *Actn1* are shown as examples. Numbers on top of the arches indicate number of reads mapping to the exon junctions. Mean of the PSI values calculated by rMATS are shown on the right. In the case of *Actn1* the PSI shown corresponds to inclusion of exon SM. Schematic of the alternative transcripts are shown below the sashimi plots, as well as the chromosome coordinates. **Bottom**, RT-PCR gel images from QIAxcel. Minus RT and no template controls were also carried out for every RT-PCR (data not shown). Values shown are mean of the PSI ± sd (n = 3). Statistical significance was calculated using Student's t-test (* $P < 0.05$, ** $P < 0.01$,*** $P < 0.001$).

Finally, we looked for any overlap between genes regulated at the mRNA abundance and AS levels (Figure 4.14). Only a few genes were found to be regulated at both levels; 41, 3 and 90 genes in the PAC1 dedifferentiation, RBPMS knockdown and RBPMS-A

overexpression respectively. Yet those numbers represent less than 4% all the genes regulated in either mRNA abundance or splicing.

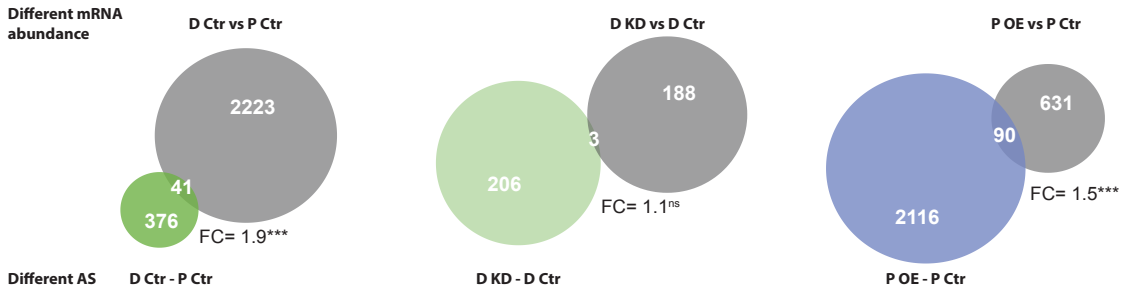


Fig. 4.14 Only a few genes are regulated at both mRNA level and splicing. Venn diagrams showing the overlap between genes displaying changes in mRNA abundance and splicing in the PAC1 dedifferentiation, RBPMS knockdown and RBPMS-A overexpression. Fold-change (FC) enrichments are shown below Venn diagrams. Significance was tested by a hypergeometric test and P-value is described as * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Therefore, the initial global data analyses confirmed RBPMS as a splicing regulator with minor effects upon mRNA abundance. The data presented so far also further support RBPMS as a master splicing regulator of the SMC differentiated program, since RBPMS promoted splicing towards the differentiated profile in the PAC1 cells.

4.2.2.3 SMC dedifferentiation ASEs regulated by RBPMS

Having obtained the PAC1 and rat aorta tissue dedifferentiation RNAseq datasets, biological comparisons could be made to address the importance of the ASE regulated in the RBPMS experiments (Figure 4.15). First, ASEs identified in PAC1 dedifferentiation and RBPMS overexpression and knockdown were compared by proportional Venn diagrams (Figure 4.15A). The diagram shows substantial overlap mainly between knockdown and PAC1 dedifferentiation datasets, representing 20% of the PAC1 regulated ASEs and ∼40% of RBPMS knockdown ASEs. To confirm that the overlapping events were reciprocally regulated, the correlation of the PSIs was verified (Figure 4.15B, left). Strikingly, all the 127 events regulated in both sets showed opposite regulation as demonstrated by the high negative correlation (r^2= -0.95). Moreover, the PAC1 dedifferentiation and overexpression also showed a reasonable correlation but positive, r^2= 0.86 (Figure 4.15B, right). Thus, the correlations between the PAC1 dedifferentiation dataset and both knockdown and overexpression experiments make biological sense. In order to better visualize those co-regulated events, 289 skipped

exons (SE) events identified in PAC1 dedifferentiation were plotted in a heatmap after performing hierarchical clustering of the samples and events (Figure 4.16). RBPMS was sufficient to reciprocally switch the splicing patterns towards the other SMC state in two identified clusters, 1 and 4, corresponding to RBPMS activated and repressed cassette exons respectively.

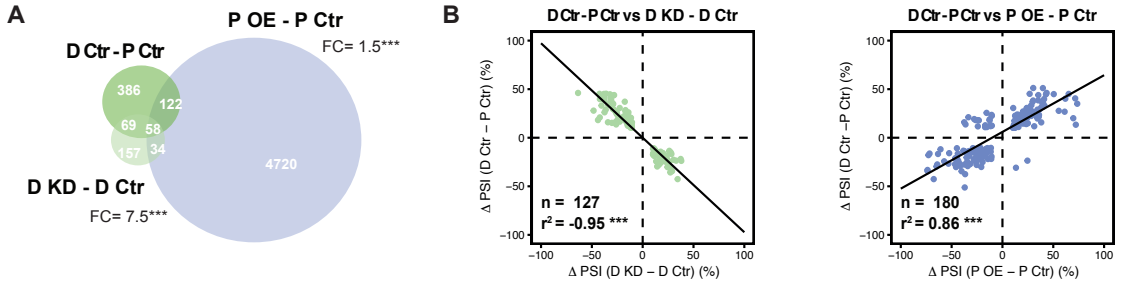


Fig. 4.15 RBPMS controls ASEs also regulated during PAC1 cells dedifferentiation. A Venn diagram of ASEs identified in RBPMS knockdown, overexpression and PAC1 dedifferentiation. Fold-change (FC) enrichment are also shown. Significance was tested by a hypergeometric test and P-value is described as * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. **B** ΔPSI correlation of the ASEs overlapping between PAC1 cells dedifferentiation and either RBPMS knockdown (green) or RBPMS-A overexpression (blue). Black line indicates the linear regression model. Pearson correlation test results are shown in the plot. The correlation statistical significance and coefficient are indicated by the p-value and r^2 in the plot.

In the same manner, comparisons were carried out using the aorta tissue dedifferentiation dataset (Figure 4.17). Of the 1714 regulated ASEs between fully differentiated aorta tissue and proliferative passage 9 (T-P9), ∼15% were also regulated by RBPMS-A overexpression in PAC1 cells (265 events). On the other hand, only a fifth of that number was found to be regulated by RBPMS knockdown (48 ASEs). The regulation of tissue-like splicing was also confirmed by the correlation of the PSI values of the overlapping ASEs, r^2= -0.39 and 0.68 for knockdown and overexpression, respectively. Additionally, in the examples shown in Figure 4.13, RBPMS-A overexpression activated and repressed the alternative exons of *Ptprf* and *Piezo1* beyond the inclusion levels observed in the differentiated PAC1. We therefore hypothesized that RBPMS overexpression could rescue fully differentiated patterns characteristic of SMC tissue-like state which are usually not seen in the PAC1 cells. Consistent with RBPMS regulation of SMC tissue-like patterns, a heatmap of tissue-regulated cassette exons (590 SE events) revealed two clusters consisting of events regulated upon RBPMS-A overexpression and during aorta dedifferentiation but not PAC1 cells nor knockdown (Figure

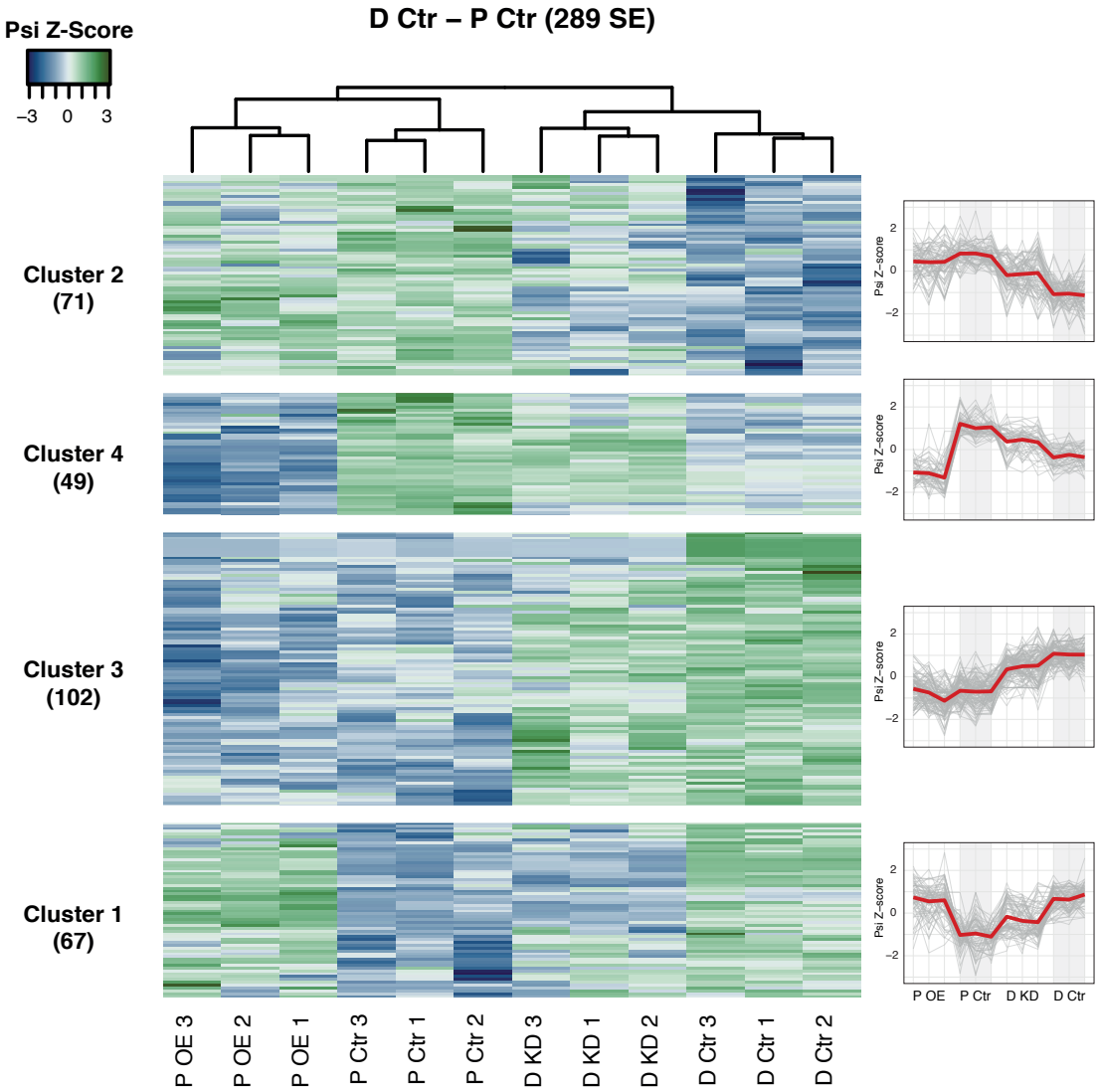


Fig. 4.16 RBPMS knockdown and overexpression reproduces PAC1 proliferative and differentiated splicing, respectively. Heatmap of PAC1 dedifferentiation regulated events (D Ctr - P Ctr comparison, 289 cassette exons). Columns represent the triplicates (1-3) from RBPMS knockdown and overexpression (D Ctr, D KD, P Ctr and P OE). Rows indicate ASEs and the blue and green colors represent low and high PSI Z-scores. Hierarchical clustering was applied to samples and ASE (columns and rows). Dendrogram illustrating hierarchical relationship between samples is shown at the top. ASE clusters are shown on the left as well as the number of events in each cluster. The PSI Z-score pattern of each cluster is plotted on the right. Mean of the PSI Z-score is shown in red.

4.18). Clusters 1 and 4, representing RBPMS activated and repressed aorta tissue SEs, accounted for ∼23% of all the SE events regulated in aorta tissue dedifferentiation. To further investigate these events in which RBPMS promoted the tissue-like state, sashimi plots were generated including RBPMS samples in PAC1 cells alongside aorta tissue dedifferentiation for comparison (Figure 4.19). *Fermt2* for example is a member of cluster 1 in which its differentiated cassette exon is only included in the aorta tissue (PSI= 36.5) and not in the PAC1 cells (PSI= ∼2.8) except when RBPMS-A is overexpressed (PSI= 38.7). Another example is *Tsc2* from cluster 4, in this case RBPMS-A overexpression repressed (PSI= 49.7) this exon which is more included in the proliferative P9 (PSI= 83.1) and not regulated in the PAC1 cells (PSI= ∼79). Lastly, *Cald1* is another example of a tissue splicing event that is never seen regulated in culture, but of which its inclusion is recapitulated by overexpression of RBPMS (∼90 folds increase in PSI). This ASE is a complex event involving a cassette exon 4 as well as an A5SS on exon 3. The use of the downstream alternative 5′SS together with exon 4 produces a transcript that encodes the heavy caldesmon isoform (h-CALD1), a well known marker of differentiated SMCs. *Cald1* AS was also validated at the protein level in western blots carried out by C. Gooding (Appendix - Figure A.2).

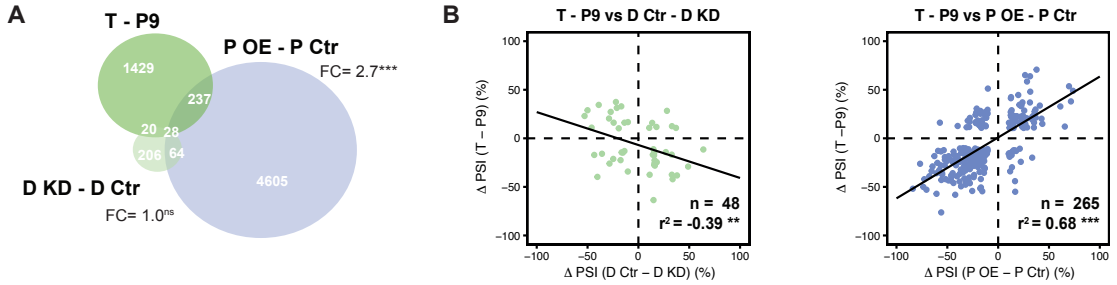


Fig. 4.17 RBPMS controls ASEs also regulated during aorta tissue dedifferentiation. A Venn diagram of ASEs identified in RBPMS knockdown, overexpression and rat aorta tissue dedifferentiation. Fold-change (FC) enrichment are also shown. Significance was tested by a hypergeometric test and P-value is described as * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. **B** ΔPSI correlation of the ASEs overlapping between rat aorta tissue dedifferentiation and either RBPMS knockdown, green, or RBPMS-A overexpression, blue. Black line indicates the linear regression model. Pearson correlation test results are shown in the plot. The correlation statistical significance and coefficient are indicated by the p-value and r^2 in the plot.

In summary, it was shown here that RBPMS can regulate splicing of a subset of SMC-specific ASEs consisting of a 20% of all the ASEs regulated during PAC1 dedifferentiation. RBPMS not only reproduced the PAC1 differentiated splicing pattern,

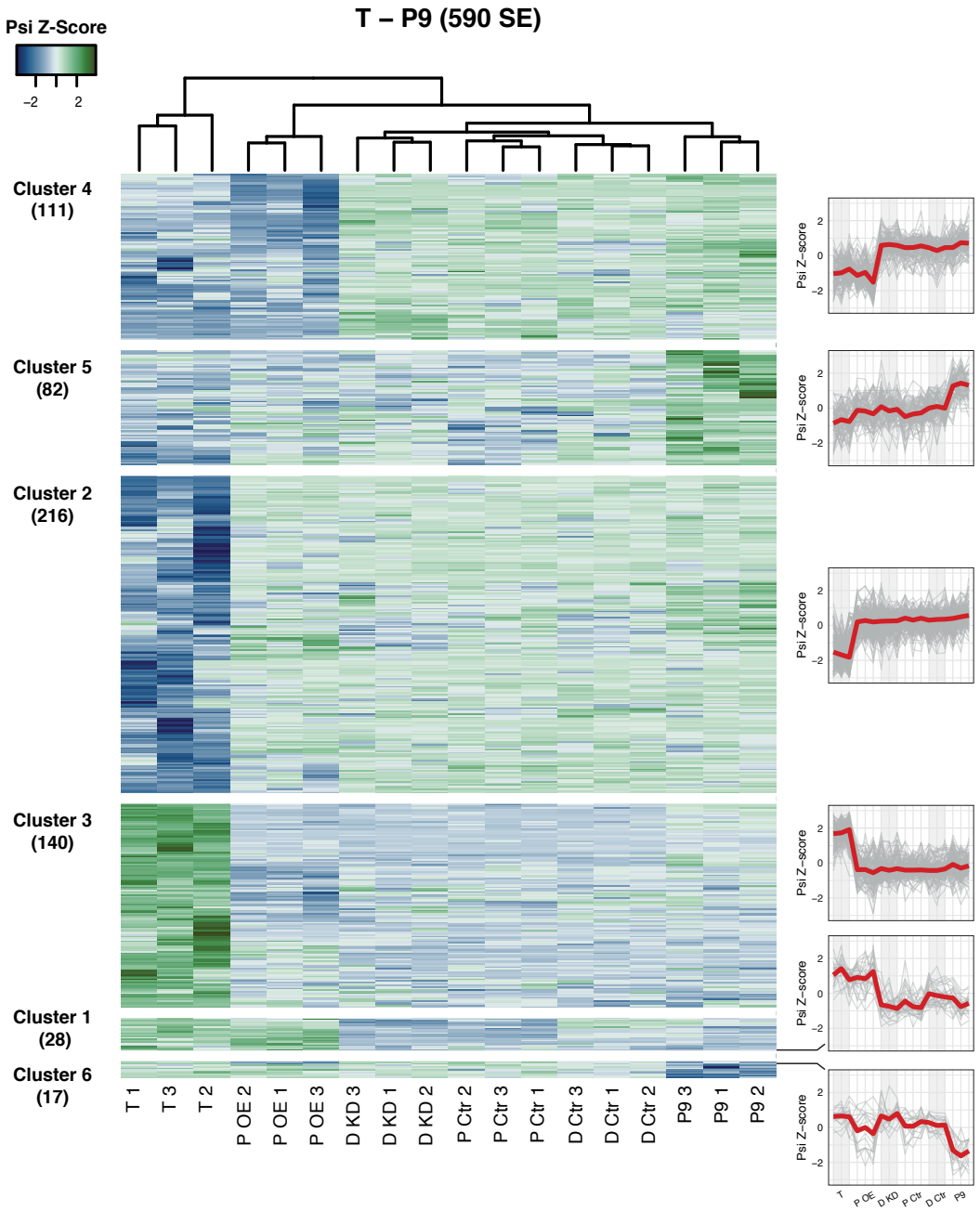


Fig. 4.18 RBPMS overexpression rescues rat aorta tissue differentiated splicing patterns. Heatmap of aorta tissue dedifferentiation regulated events (T - P9 Ctr comparison, 590 cassette exons). Columns represent the triplicates (1-3) from aorta dedifferentiation, RBPMS knockdown and overexpression (T, P9, D Ctr, D KD, P Ctr and P OE). Rows indicate ASEs and the blue and green colors represent low and high PSI Z-scores. Hierarchical clustering was applied to samples and ASE (columns and rows). Dendrogram illustrating hierarchical relationship between samples is shown at the top. ASE clusters are shown on the left as well as the number of events in each cluster. The PSI Z-score pattern of each cluster is plotted on the right. Mean of the PSI Z-score is shown in red.

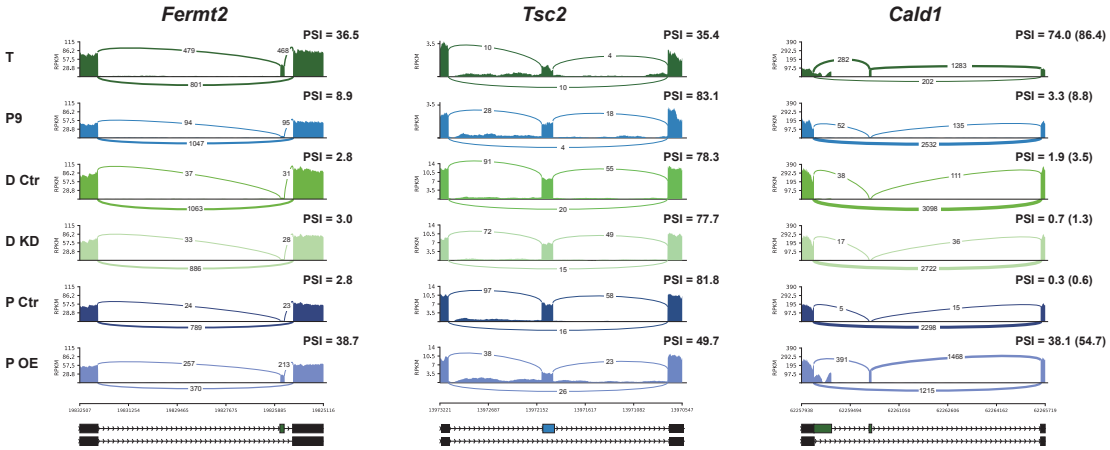


Fig. 4.19 RBPMS strongly regulates tissue patterns of *Fermt2*, *Tsc2* and *Cald1* Sashimi plots of ASEs in which RBPMS overexpression recapitulates patterns only seen in tissue. *Fermt2* from Cluster 1, and *Cald1* mis-clustered to Cluster 3 are examples of RBPMS activation whereas *Tsc2* represents a repression event from Cluster 4. Numbers on top of the arches indicate number of reads mapping to the exon junctions. Means of the PSI values calculated by rMATS are shown on the right. For *Cald1*, an alternative PSI was manually calculated to account its A5SS and is shown in front of the rMATS PSI. Schematic of the alternative transcripts are shown below the sashimi plots, are the chromosome coordinates.

but also recapitulated tissue-like splicing patterns that are rarely observed in cultured cells.

4.2.2.4 RBPMS global target functional analysis

To gain more information about the genes regulated by RBPMS, Gene Ontology (GO) analyses were carried out as described in the Material and Methods chapter. Genes with changes in their transcript levels upon RBPMS manipulation either did not show enrichment or were enriched for functions involved in stress response (Tables 4.5 and 4.6). On the other hand, the PAC1 dataset was enriched for functions involved in the SMC phenotypic switch, such as muscle development, contraction, cell proliferation and migration (Table 4.7).

Table 4.5 GO analysis of genes differentially abundant upon RBPMS knockdown

GO.Term	Description	Enrichment	N	Category
Downregulated				
No enrichment				CC
No enrichment				MF
No enrichment				BP
Upregulated				
GO:0005615	Extracellular space	3.5	16	CC
GO:0005576	Extracellular region	4.6	11	CC
No enrichment				MF
GO:0051707	Response to other organism	7.0	20	BP
GO:0009607	Response to biotic stimulus	5.1	25	BP
GO:0043207	Response to external biotic stimulus	5.17	24	BP
GO:0009615	Response to virus	9.0	13	BP
GO:0006952	Defense response	5.0	19	BP

Next, GO enrichment was assessed for the genes whose splicing was affected by RBPMS knockdown, overexpression or during PAC1 and aorta dedifferentiation (Figure 4.20). RBPMS knockdown targeted genes involved in processes, components and functions similar to the SMC dedifferentiation (PAC1 and aorta), for instance actin filament based process, cell junction and regulation of GTPase (Appendix - Table A1, for list of genes from enriched GO terms). RBPMS-A overexpression affected splicing of genes associated with regulation of GTPase activity and microtubule organization which are relevant to SMC biology, however it also affected genes with various other functions with less clear implications on the SMC biology.

In view of the fact that genes marked by super-enhancers are important for the cell function and identity, it was tested whether RBPMS knockdown-mediated ASEs were enriched for SM super-enhancer associated genes (Figure 4.21). Enrichment was verified for three different SM tissues super-enhancers (aorta, bladder and stomach smooth

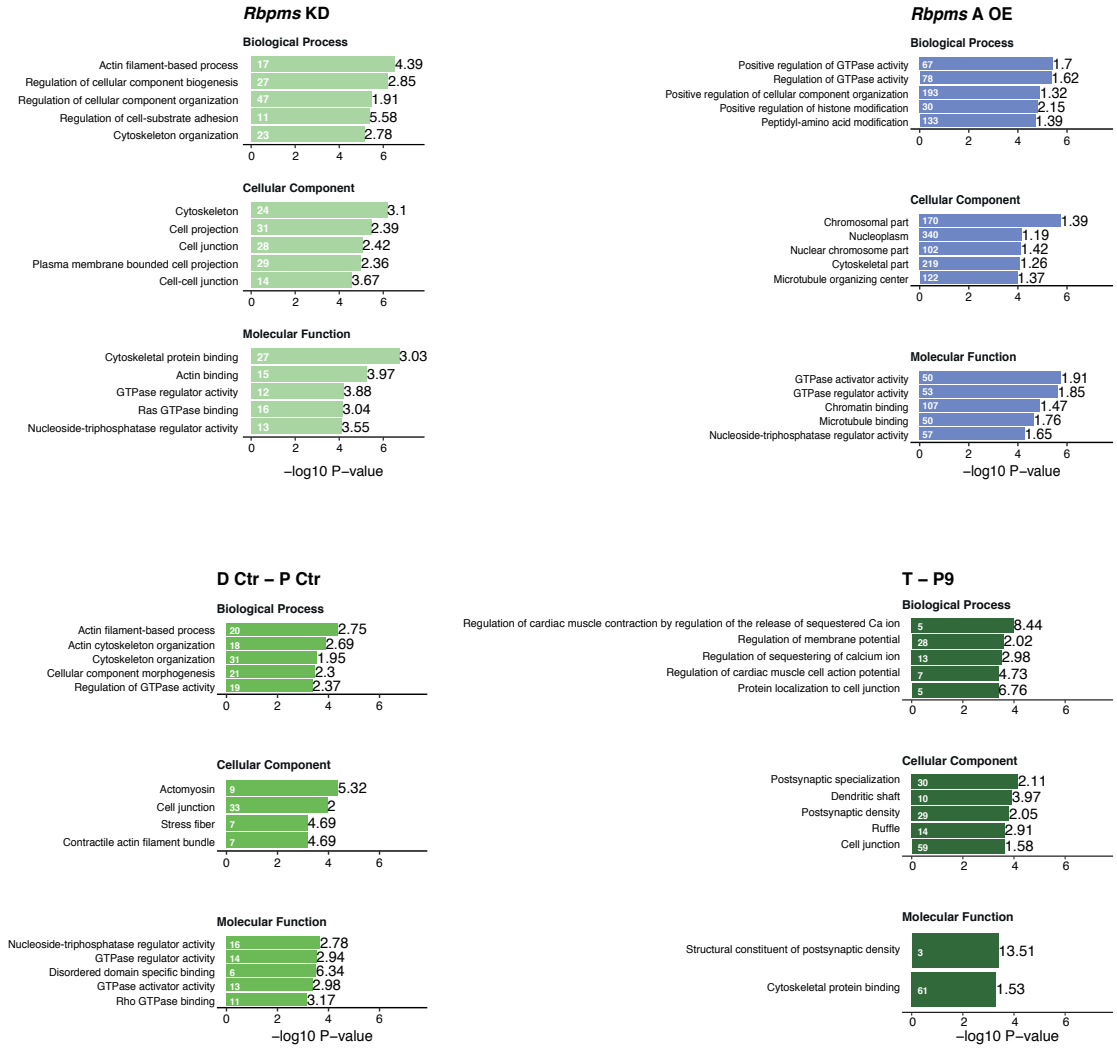


Fig. 4.20 RBPMS regulated ASE are enriched for SMC functions. GO analysis of genes differentially spliced by RBPMS knockdown (KD), overexpression (OE), PAC1 dedifferentiation (D Ctr - P Ctr) and aorta tissue dedifferentiation (T - P9). Enriched GO terms were ranked by p-value and only the top five are shown for each category. Numbers in front and within the bars indicate the fold enrichment relative to the background number of genes and the number of genes in the enriched term.

Table 4.6 GO analysis of genes differentially abundant upon RBPMS-A overexpression

GO.Term	Description	Enrichment	N	Category
Downregulated				
No enrichment				CC
No enrichment				MF
GO:0006575	Cellular modified amino acid metabolic process	8.42	11	BP
Upregulated				
GO:0005887	Integral component of plasma membrane	2.3	37	CC
GO:0005615	Extracellular space	2.1	45	CC
GO:0044421	Extracellular region part	1.9	56	CC
GO:0031226	Intrinsic component of plasma membrane	2.2	39	CC
GO:0099056	Integral component of presynaptic membrane	5.6	10	CC
GO:0004888	Transmembrane signaling receptor activity	3.0	29	MF
GO:0038023	Signaling receptor activity	2.6	35	MF
GO:0005230	Extracellular ligand-gated ion channel activity	14.9	6	MF
GO:0060089	Molecular transducer activity	2.4	37	MF
GO:0005231	Excitatory extracellular ligand-gated ion channel activity	18.7	5	MF
GO:0006952	Defense response	3.1	55	BP
GO:0098542	Defense response to other organism	4.1	32	BP
GO:0051707	Response to other organism	3.2	43	BP
GO:0051607	Defense response to virus	4.86	20	BP
GO:0009615	Response to virus	3.8	25	BP

muscle), revealing significant enrichment for genes associated with super-enhancers in all of them. Approximately 15% of the splicing events were in super-enhancer associated genes with an average of 2 fold increase in the enrichment compared to the background. Thus, these data further enhance the importance of the RBPMS-affected genes to the SMC biology.

Finally, the potential protein-protein interaction network affected by RBPMS regulation of splicing was assessed using STRING (Szklarczyk et al., 2017) (Figure 4.22). For that, a more restricted list of ASE was created by combining the RBPMS knockdown and PAC1 dedifferentiation co-regulated ASEs with the events commonly regulated between RBPMS overexpression and aorta tissue (Material and Methods Chapter). The PPI network obtained for RBPMS regulated ASEs showed genes involved in protein phosphorylation (red), cell-substrate adherens junctions (yellow) and cytoskeletal protein binding (blue). Many of the ASEs strongly regulated by RBPMS, e.g. ACTN1, FLNB and ITGA7 (Table 4.3), are found as hubs in the network. Furthermore, the functional importance of several of the nodes in the PPI network was reinforced by their association with SM super-enhancers as highlighted in green (Figure 4.22).

Therefore, in agreement with the hypothesis of RBPMS being a master regulator of SMC, its targets comprised a functionally coherent set of genes involved in critical SMC functions that were also associated with super-enhancers themselves.

Table 4.7 GO analysis of genes differentially abundant upon PAC1 dedifferentiation

GO Term	Description	Enrichment	N	Category
Downregulated				
GO:0005887	Integral component of plasma membrane	3.0	140	CC
GO:0044459	Plasma membrane part	2.0	293	CC
GO:0031226	Intrinsic component of plasma membrane	2.9	145	CC
GO:0031224	Intrinsic component of membrane	1.6	430	CC
GO:0016021	Integral component of membrane	1.6	416	CC
GO:0004888	Transmembrane signaling receptor activity	3.2	113	MF
GO:0060089	Molecular transducer activity	2.7	89	MF
GO:0004930	G protein-coupled receptor activity	3.9	118	MF
GO:0005216	Ion channel activity	2.8	42	MF
GO:0022838	substrate-specific channel activity	2.83	50	MF
GO:0007165	Signal transduction	1.5	364	BP
GO:0032501	Multicellular organismal process	1.6	304	BP
GO:0051239	Regulation of multicellular organismal process	1.5	345	BP
GO:0007166	Cell surface receptor signaling pathway	1.7	197	BP
GO:0032502	Developmental process	1.3	489	BP
Upregulated				
GO:0031224	Intrinsic component of membrane	1.7	234	CC
GO:0016021	Integral component of membrane	1.7	227	CC
GO:0044459	Plasma membrane part	2.0	152	CC
GO:0044425	Membrane part	1.5	303	CC
GO:0044421	Extracellular region part	2.2	96	CC
GO:0004888	Transmembrane signaling receptor activity	3.2	46	MF
GO:0038023	Signaling receptor activity	2.7	55	MF
GO:0060089	Molecular transducer activity	2.5	57	MF
GO:0048018	Receptor ligand activity	3.0	28	MF
GO:0004930	G protein-coupled receptor activity	3.8	20	MF
GO:0007165	Signal transduction	1.5	196	BP
GO:0051239	Regulation of multicellular organismal process	1.5	178	BP
GO:0007166	Cell surface receptor signaling pathway	1.7	101	BP
GO:0009582	Detection of abiotic stimulus	5.6	14	BP
GO:0050920	Regulation of chemotaxis	3.2	27	BP

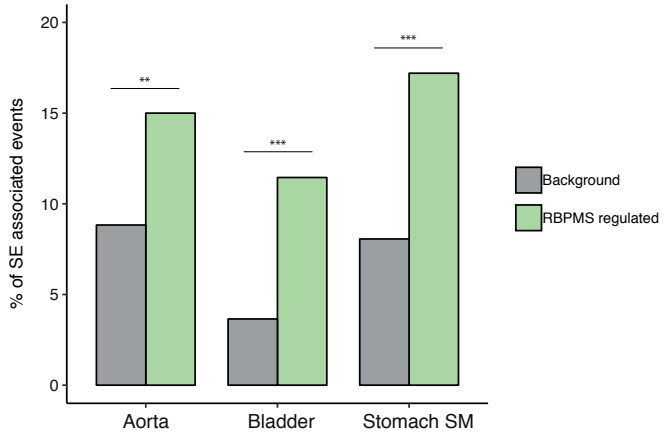


Fig. 4.21 RBPMS regulated ASE are enriched in genes associated with SM super-enhancers. Enrichment for SM super-enhancer associated genes in the RBPMS knockdown regulated ASE. The set of all cassette exons detected by rMATS (unregulated and regulated) in RBPMS knockdown is shown as background. Significance was tested by a hypergeometric test and P-value is described as * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

4.3 Discussion

4.3.1 RBPMS is an AS regulator

RBPMS and RBPMS2 are conserved across vertebrates with orthologs present in the animal models *Drosophila* and *C. elegans*, known as Couch Potato and MEC-8 respectively (Soufari and Mackereth, 2017). Although RBPMS and RBPMS2 proteins localize to both nucleus and cytoplasm, it is mainly their cytoplasmic roles that have been investigated. The nuclear role of RBPMS is restricted to transcriptional co-regulation (Fu et al., 2015; Sun et al., 2006) whereas its cytoplasmic roles are involved in several RNA biology processes such as mRNA stability (Rambout et al., 2016), transport (Hörnberg et al., 2013) and localization in cytoplasmic granules (Farazi et al., 2014; Furukawa et al., 2015; Hörnberg et al., 2013). RBPMS2 was shown to interact with eEF2, eukaryote elongation factor-2, to regulate translational control of specific RNAs of gastrointestinal SMC (Sagnol et al., 2014).

RBPMS has been reported to act as a transcriptional co-activator by enhancing the transcription activity of SMAD proteins and also by repressing a family of transcriptional factors, AP-1 (Fu et al., 2015; Sun et al., 2006). However, reciprocal changes in mRNA abundance upon manipulation of RBPMS levels were restricted to only four genes other than *Rbpms* (*Fst*, *P4ha2*, *AABR07022144.1*, *LOC100911692*), providing no

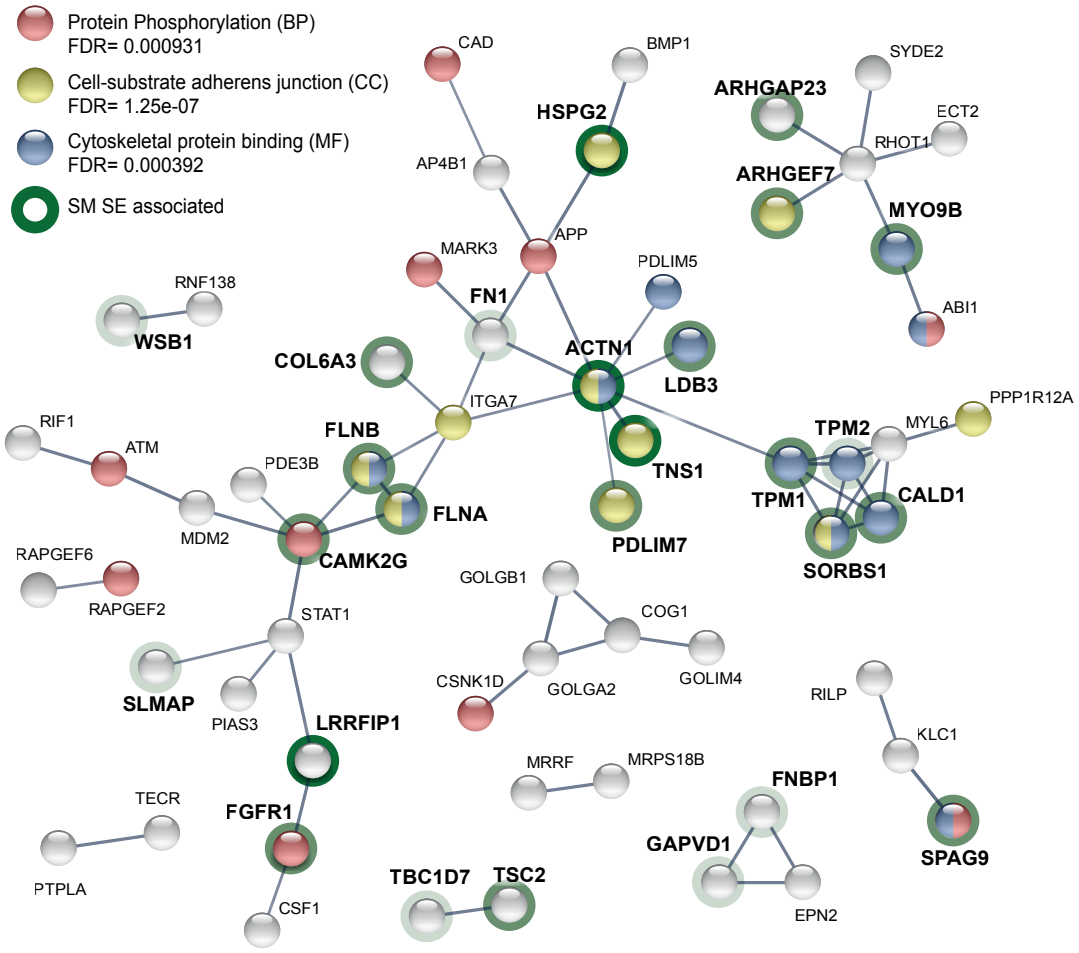


Fig. 4.22 RBPMS affects a PPI network involved in SMC functions. Protein-Protein Interaction (PPI) network was generated using STRING and a set of ASEs from the combination of the overlap between RBPMS knockdown and PAC1 dedifferentiation and the overlap between RBPMS-A overexpression and aorta dedifferentiation. Network edges indicate interaction confidence. GO terms enriched in the dataset are also highlighted in red, blue and yellow. Genes associated with SM super-enhancers are shown in bold font and are also highlighted in green according to the number of SM tissues where they were found associated with super-enhancers (aorta, bladder or stomach SM), dark green represents association in all the SM tissues.

evidence for a transcriptional role by RBPMS. In agreement with the SMC literature, PAC1 cells dedifferentiation was accompanied by substantial changes in the mRNA abundance of many genes, such as SMC differentiation markers (Owens et al., 2004; Rothman et al., 1992). In terms of splicing changes, both RBPMS knockdown and overexpression dramatically affected AS of the PAC1 cell transcriptome. In fact, changes in mRNA abundance upon RBPMS were largely outnumbered by regulated AS events. Furthermore, other studies disclosed the remarkable feature that transcriptional and post-transcriptional controls tend to focus on different set of genes. Therefore it is not surprising that little overlap was observed between the genes controlled at the expression and AS for all the comparisons analyzed here (Blencowe, 2006).

RBPMS family members, apart from MEC-8, have not been implicated in splicing regulation. In the nematode *C. elegans*, RBPMS homolog, MEC-8, was described to regulate splicing of *Unc-52* transcripts (Lundquist et al., 1996). In the only study that focused on RBPMS functions in mRNA processing in the human HEK293 cell line PAR-CLIP and mRNA-seq were performed for the identification of RBPMS preferred sites and transcriptome-wide targets, but no regulated ASEs associated with RBPMS binding was identified (Farazi et al., 2014). Nevertheless, reanalysis of RBPMS overexpression in HEK293 from Farazi et al. (2014) using rMATS was performed by Dr. Miriam Llorian and it revealed a few regulated ASEs, including *ITGA7*, *FLNB* (Figure 4.23) and *FN1* also identified in the PAC1 cells. Actually, as a splicing regulator RBPMS can only regulate actively transcribed genes, thus it is unsurprising that only a few genes were found differentially spliced in HEK293 cells, given that RBPMS mainly regulated SMC transcripts. In that way, it seems unlikely that RBPMS is sufficient to initiate SMC differentiation.

Finally, although related family members are sometimes able to compensate for the manipulation of an RBP level (Mockenhaupt and Makeyev, 2015), no alterations in *Rbpms2* transcript levels were observed in any of the treatments. Additionally, *Rbpms2* knockdown alone had no effects on splicing and when combined with *Rbpms* knockdown it did not increase the splicing changes. This is probably explained by the low expression levels of the paralog in the SMCs (Figure 3.4 and 4.7).

Therefore, the data presented here provide the strongest evidence to date for the widespread molecular function of RBPMS as a splicing regulator.

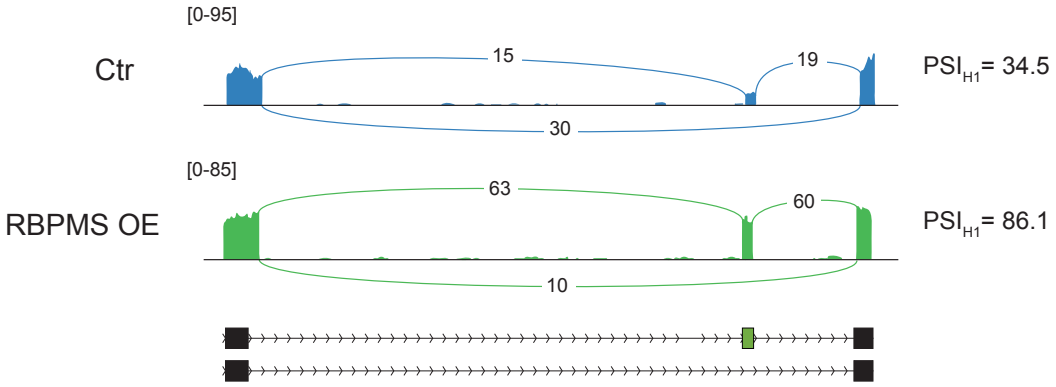


Fig. 4.23 FLNB alternative splicing event found by reanalysis of Farazi *et al* 2014 using rMATS. Sashimi plot for *Flnb* exon H1, identified in the reanalysis of Farazi *et al* 2014. Numbers on top of the arches indicate number of reads mapping to the exon junctions. Mean of the PSI values calculated by rMATS are shown on the right. Schematics of the transcript isoforms are shown at the bottom. Regulated H1 exon is highlighted in green.

4.3.2 RBPMS promotes the differentiated AS program of VSMCs

RNAseq of PAC1 cells in which RBPMS levels were manipulated revealed this RNA-binding protein (RBP) as a splicing regulator in SMCs, repressing and activating hundreds of ASEs (Figure 4.11), majority of which were regulated towards a more differentiated SMC program, as indicated by the correlation coefficients (Figure 4.15). Indeed, RBPMS was found associated with super-enhancers in adult human SM tissues (Hnisz et al., 2013) and besides promoting the SMC patterns in the PAC1 cells, RBPMS overexpression was able to reproduce aorta tissue splicing patterns in the proliferative PAC1 cells (Figure 4.18 and 4.19). *Cald1*, *Fermt2*, *Tsc2*, *Tns1*, *Tpm1* and *Actn1* were some of the ASEs whose splicing levels were comparable to fully differentiated tissue SMC. In the case of *Cald1*, *Fermt2*, *Tsc2* and *Tns1* their regulation was never seen in the cultured PAC1 cells; but RBPMS was sufficient to promote their fully differentiated tissue-like splicing. Thereby, this suggests that RBPMS may play a key role in maintaining adult SMC phenotype. In addition, several studies have reported RBPMS anti-proliferative tumor-suppressive activity, supporting its role in the non-proliferative SMC state (Fu et al., 2015; Hou et al., 2018; Rastgoo et al., 2018).

Despite the lack of RBPMS studies in SMCs, its paralog RBPMS2 has been implicated in a critical role in gastrointestinal SMCs. RBPMS2 is indeed found to be associated with super-enhancers in stomach SM and it was shown to be expressed in visceral SMC during early developmental stages, decreasing in mature cells (Notarnicola et al., 2012). In contrast with the hypothesis suggested here and with the results observed upon manipulation of RBPMS levels in PAC1 cells, sustained overexpression of RBPMS2 in chicken visceral SMCs induced their dedifferentiation and loss of contractility (Notarnicola et al., 2012). RBPMS2 stimulated the proliferative phenotype by increasing *Noggin* expression as well as translationally upregulating Noggin via interaction with an elongation factor, eEF2 (Notarnicola et al., 2012; Sagnol et al., 2014). Noggin then leads to inhibition of the BMP signaling, a pathway that prevents SMC dedifferentiation (Notarnicola et al., 2012; Sagnol et al., 2014). However, no regulation of *Noggin* transcripts was observed in any of the PAC1 conditions analyzed by mRNA-seq. Indeed its expression was only detected in the proliferative cells at a very low level ($\sim$3 TPM). Finally, similar to RBPMS-A, RBPMS2 promoted the SMC splicing pattern when overexpressed in PAC1, HeLa and HEK293 cells (Figure 3.12, 3.14, 3.13), indicating that RBPMS paralogs have an intrinsically similar molecular activity. The reasons determining the discrepancy between the promotion of differentiated splicing patterns by RBPMS2 but the dedifferentiation phenotype reported in visceral SMCs remain unclear. Variation of pre-mRNAs and mRNA targets, subcellular localization, interacting protein partners, post-translational modifications and in cell signaling pathways are all plausible explanations for the differential modulation of RBPMS and RBPMS2 activity across SM tissues.

Thus, in VSMC RBPMS acts as a splicing regulator by activating and repressing splicing of ASEs towards the mature differentiated SM AS program. Moreover, previous results suggest this to be a feature shared between the paralogs.

4.3.3 RBPMS regulates AS of mRNAs important for SMC functions

SMC phenotypic plasticity is marked by the capability of interconversion between a more contractile state and proliferative, migratory and secretory phenotype (Owens et al., 2004). This switching is concerted by the modulation of the repertoire of contractile proteins as well as its rearrangement in order to confer more motility to individual cells instead of composing the tissue-contraction machinery (Ye et al., 2014). This remodeling is achieved by regulation of the actin cytoskeleton network and by its

anchoring to integrins in focal adhesions, which then determines the connections to the extracellular matrix (ECM). Furthermore, this reorganization of contractile and cytoskeletal proteins has been reported to also happen to aortic SMC in culture (Worth et al., 2001). Strikingly, RBPMS-regulated AS of transcripts that are involved in the cytoskeleton organization process, including targets with functions in cytoskeleton binding and more specifically actin binding (Figure 4.20 and 4.22). Additionally, the integrin adhesome (Horton et al., 2015), ECM components and GTPase activity regulators were other targets of RBPMS-mediated AS. These functions are all critical for the cytoskeleton reorganization by contributing to the cell-matrix adhesion and cytoskeleton dynamics (Bar-Sagi and Hall, 2000; Frismantiene et al., 2018; Horton et al., 2015). Consistent with that, both PAC1 cells and aorta tissue dedifferentiation exhibited similar GO terms enrichment to the RBPMS-regulated AS events. Further evidence of the importance of RBPMS targets was provided by the fact that they also show proximity to super-enhancers in SM tissues (Figure 4.21). Actually, some of these, for instance *Actn1*, *Flnb* and *Tns1*, are super-enhancer associated, directly interact with actin and integrins and constitute three of the four modules defined by the consensus integrin adhesome (Horton et al., 2015). In fact, *Vcl*, which was also found to be regulated by RBPMS overexpression, is a major component of the fourth module, establishing interactions to components of the other modules. However, its interaction to integrin is indirect and mediated by Talin proteins (Horton et al., 2015).

Focal adhesions and the integrin adhesome comprise multi-molecular mechanosensitive complexes, participating in both adhesion to the ECM as well as intracellular signaling by sensing mechanical cues from the environment (Wehrle-Haller, 2012). The adhesion dynamics are in part controlled by phosphorylation of tyrosine residues of its components, such as PXN and FAK (Wehrle-Haller, 2012). Interestingly, a small cluster formed by the receptor tyrosine phosphatase PTPRF and two interacting proteins PPFIA1 and PPFIBP1 are found regulated by RBPMS (Figure 4.12 and 4.13). RBPMS promoted inclusion of the LAR Alternatively Spliced Element-a (LASE-a) in the *Ptprf* transcripts. This exon adds a short peptide sequence, GSSAPSCPNISS, to the proximal membrane region of the PTPRF protein (Honkaniemi et al., 1998). This event has been associated with intraneuronal localization via potential targeting of the introduced serine residues to phosphorylation (Honkaniemi et al., 1998). In contrast, the roles of the interacting proteins splicing isoforms, PPFIA1 and PPFIBP1, are less understood. Another regulator of cell adhesion found differentially spliced upon RBPMS is the mechanosensitive ion channel coding gene *Piezo1* (Figure 4.13). PIEZO1 is critical in SMCs during arterial remodeling in hypertension (Retailleau et al., 2015)

and although functional differences of its paralog, PIEZO2, splicing isoforms have been reported (Szczot et al., 2017), not much is known about the RBPMS-repressed exon. The regulated event lies within the conserved Piezo1 domain, neighboring the mechanosensing "beam" (Liang and Howard, 2018). Therefore, it could probably lead to effects on its protein function.

Components of the ECM were also affected by RBPMS, including FN1, COL6A3 and HSPG2. Fibronectin (FN1) exon EDB (also referred as EDII) was more included by RBPMS. Functional consequences are restricted to *in vitro* studies showing mild defects in matrix assembly and proliferation by $EDB^{-/-}$ embryonic fibroblasts (MEF) (White et al., 2008). However, no significant phenotype was observed *in vivo* $EDB^{-/-}$ mice (White et al., 2008). The COL6A3 splicing isoforms have not yet been functionally characterized, although these ASEs were reported to be affected in colon cancer (Gardina et al., 2006). Notably, *Hspg2* encodes the perlecan protein, a key component of the vascular ECM, which is actually the identified splicing target of MEC-8 in *C. elegans* (Lundquist et al., 1996).

Interestingly a small network identified in the PPI analysis, involves proteins from the secretory pathway, GOLIM4, GOLGA2, GOLGB1 and COG1. This reflects the transition from a differentiated phenotype to a more synthetically active cell state observed in the proliferative SMC (Owens et al., 2004). Indeed, a recent genome-wide study identified AS as a regulator of the protein transport efficiency by generation of alternative isoforms in a tissue-specific manner (Neumann et al., 2019). Therefore, it is possible that regulation of these events by RBPMS could have direct effects in the SMC secretory pathway, yet more functional studies are required to elucidate the role of these isoforms in the SMC biology.

In a previous study from this laboratory, the AS program of the differentiated SMCs was for the first time profiled using an exon-junction microarray (Llorian et al., 2016). Dedifferentiation of mouse SM tissues (aorta and bladder) was shown to be achieved by concerted non-productive splicing of post-transcriptional regulators, involving both auxiliary RBP and core splicing machinery components (Llorian et al., 2016). However, significant changes in the expression of these factors were not observed in the PAC1 dedifferentiation or upon RBPMS manipulation levels. This observation might be explained by the expected differences between *in vivo* and *in vitro* SMC cultures, for instance tissue dedifferentiation led to 5660 differential expressed genes whereas approximately only half of that was regulated in the PAC1 cells (2774). Aorta dedifferentiation also promoted far more alternative splicing changes (1714 events) than PAC1 cells (635). Therefore, although retaining differentiated properties that

allow to interrogate SMC dedifferentiation in culture (Rothman et al., 1992), the PAC1 cell line has its limitations.

Thus, consistent with the hypothesis that RBPMS is a master regulator of AS in SMCs, a functionally coherent set of genes important for SMC function was affected by RBPMS mediated splicing. Although the consequences of some of the splicing isoforms promoted by RBPMS are better understood, the roles of many other regulated ASEs remain to be characterized.

4.4 Final conclusions

In summary, this chapter aimed to uncover the global changes in AS upon RBPMS knockdown and overexpression using a high-throughput RNAseq approach. Profiling of the transcriptome of PAC1 cells upon manipulation of RBPMS levels allowed the discovery of numerous RBPMS-regulated ASEs. RBPMS affected all five categories of AS by activating and repressing splicing. Moreover, RBPMS controlled splicing of 20% of the events regulated during the PAC1 dedifferentiation (50% of RBPMS knockdown ASEs) and RBPMS levels are sufficient to explain these splicing changes during this transition. Surprisingly, RBPMS overexpression was also capable of fully recapitulating some SM tissue splicing patterns which are not seen in the cultured PAC1 cell line. Thus, RBPMS not only promoted the differentiated PAC1 SMC pattern but also established a further differentiated profile typical of a SM tissue state. Consistent with the master regulator hypothesis, RBPMS targeted a functionally related subset of mRNAs involved in SMC functions, for instance cytoskeleton, GTPase activity and cell adhesion. These are all functions important for the SMC switching from a contractile differentiated to a more motile and proliferative state. Moreover, the PPI network of RBPMS targets also revealed a cluster of genes involved in the secretory pathway, a feature critical for the more proliferative SMCs that are more synthetically active. Therefore, these data established RBPMS as a splicing regulator in VSMCs, promoting the differentiated AS program in these cells as expected from a master splicing regulator.

In conclusion, this chapter has the following points as the main findings:

1. RBPMS knockdown and overexpression of RBPMS-A alter AS of PAC1 cells.
2. RBPMS affected all AS types, with no preference towards any category.
3. RBPMS was able to repress and activate splicing, displaying some bias towards repression in the case of RBPMS-A overexpression.

4. RBPMS effects on AS outnumber mRNA abundance changes, with little overlap between genes regulated by both processes.
5. RNAseq analysis by rMATS detected AS changes with high confidence as experimentally validated by RT-PCR.
6. RNAseq allowed the profiling of the differentiated and proliferative PAC1 cell transcriptomes, providing insights on the mRNA changes (abundance and AS) during the phenotypic switching of this SMC cell line.
7. RBPMS promotes the SMC splicing pattern; accounting for 20% of all the AS changes observed during PAC1 cells dedifferentiation.
8. RBPMS-A overexpression reproduced the adult aorta tissue splicing pattern (e.g. *Cald1*, *Fermt2* and *Tsc2*), not usually observed in the PAC1 cell line.
9. RBPMS affected splicing of transcripts involved in SMC functions as indicated by the GO analysis.
10. RBPMS could potentially disrupt a PPI network associated with the SMC transition from a contractile to a motile phenotype by rearrangement of the cytoskeleton components.
11. RBPMS targets are enriched for SM super-enhancer associated genes, reinforcing their importance to the SMC biology.

In addition to that, this chapter also led to other questions listed below:

- Does RBPMS regulation of the SMC splicing lead to phenotypic changes?
- Does RBPMS regulation of the cytoskeleton genes affect SMC contractility or motility?
- **Does RBPMS repress the exon NM or activate exon SM in *Actn1* splicing?**
- **Does RBPMS directly regulate splicing of these targets?**
- **Are the conserved CACs important for RBPMS regulation of the SMC splicing?**
- **What is the mechanism behind RBPMS regulation of splicing?**
- **How do the isoforms display different activities in splicing?**

The questions in bold were addressed in the following chapters of this study.

Chapter 5

RBPMS: a direct regulator of alternative splicing

5.1 Introduction

5.1.1 RBPs in the regulation of AS

5.1.1.1 Direct regulation of AS

AS is primarily regulated by the combination of *cis*-elements embedded in the pre-mRNA sequence and *trans*-factors comprised of RBPs (reviewed in (Fu and Ares, 2014)). The regulatory sequences act as splicing enhancers or silencers depending on the context that they are found. This reflects in the position-dependent principles that underlie the regulation of AS by RBPs, a feature outlined by several biochemical and molecular studies and better understood now with more global approaches (Witten and Ule, 2011). Generally, splicing regulatory RBPs are able to repress splicing upon interaction with exonic and upstream intronic regions and to activate inclusion upon binding to sites in the downstream intron (Fu and Ares, 2014). Therefore, these common set of rules followed by different RBPs indicate shared mechanisms for regulating the splicing apparatus. Eventually, further studies of other RBPs and RNA splicing maps will shed more light on the global principles of the position-based splicing and identify other specific mechanisms that differ from those proposed so far.

5.1.1.2 Indirect regulation of AS

Aside from the direct regulation of AS, RBPs are known to regulate indirect targets as a result of cross-regulation of other splicing regulators or via protein-protein interactions

that allow RNA interaction (Fu and Ares, 2014; Jangi and Sharp, 2014). The former is achieved by the regulation of AS-NMD events within RBPs or by production of protein isoforms that are either nonfunctional or dominant-negative (Jangi and Sharp, 2014). A well characterized example of cross-regulation between post-transcriptional factors is the PTBP1/PTBP2 switch in neuronal cells. PTBP1 regulates an event in PTBP2 that when skipped triggers NMD (Boutz et al., 2007a,b; Makeyev et al., 2007; Spellman et al., 2007). However, during neuronal differentiation, downregulation of PTBP1 by miR-124 relieves its repressive activity upon PTBP2 allowing expression of PTBP2 in these cells (Makeyev et al., 2007). Nowadays, the advances in global transcriptome approaches such as RNAseq and iCLIP (Individual-nucleotide resolution Cross-Linking and ImmunoPrecipitation) allow distinguishing directly regulated targets and uncovering events that are actually part of a cascade of secondary splicing changes due to direct regulation of splicing factors (Wollerton et al., 2004). For instance, by applying these two approaches in parallel, Rbfox2 was shown to affect the gene expression of a network of RBPs via AS-NMD subsequently causing indirect splicing changes (Jangi et al., 2014). Moreover, other RBPs are expected to be regulated in a similar manner, since genes encoding splicing factors are enriched for AS-NMD events (Jangi and Sharp, 2014). Finally, RBPs are likely to establish protein-protein interactions and whether homotypic or heterotypic they all could affect the splicing outcome (Witten and Ule, 2011). AS decisions can be modulated by the different conformational changes in the pre-mRNA depending on the RBPs bound to it. Alternatively, larger complexes comprising different RBPs can expand the targets of an RBP by allowing binding to motifs that are specific to other components of the complex. Indeed, the latter has been shown to be the case for RBFOX proteins which take part in a large assembly of splicing regulators (LASR) involving several hnRNPs and other RBPs. Interestingly, RBFOX was able to affect splicing using LASR component hnRNP M binding sites (Damianov et al., 2016).

Therefore, RBPs are critical regulators of AS, affecting splicing in a direct and indirect manner. In addition, the combinatorial nature of AS regulation is also important for the establishment of tissue-specific programs (Baralle and Giudice, 2017; Jangi and Sharp, 2014).

5.1.2 RNA binding by the RBPMS RRM domain

PAR-CLIP combined with structural analysis of RBPMS uncovered its RNA binding features (Farazi et al., 2014; Sagnol et al., 2014; Teplova et al., 2016). RBPMS binding to RNA was observed in several studies, but it was only in Farazi et al. (2014) that

RBPMS global RNA targets were determined in HEK293, leading to the identification of its binding motif as tandem CAC trinucleotides separated by variable spacer length (1-10 nt). This recognition element has also been shown to be the preferred binding site of the RBPMS ortholog proteins, CPO and MEC-8 in *C. elegans* and *Drosophila* respectively (Soufari and Mackereth, 2017). PAR-CLIP defined RBPMS-binding sites were distributed among 3′-UTR, intron, exons and repetitive regions. More than half of the binding sites were found in the former two locations (Farazi et al., 2014). Moreover, the crystal structure of human RBPMS RRM (residues 14-111), both free and bound to UCACU RNA molecules, allowed further characterization of RBPMS residues involved in RNA binding (Teplova et al., 2016). Briefly, RBPMS RRM comprises of four-stranded anti-parallel β-sheets packed against a pair of α-helices, thus assuming the classical RRM fold. Conserved aromatic amino acids, Phe27 and Phe65, were key residues in the establishment of the intermolecular interaction. Other residues involved in the RNA binding are highlighted in Figure 5.1. Consistent with the pair of CACs identified by PAR-CLIP as the RBPMS RRM recognition element, structural studies revealed RBPMS dimerisation properties (Teplova et al., 2016). Dimerisation was achieved by residues within the RRM domain and it was critical to RBPMS RNA binding. RBPMS dimeric nature is further discussed in Chapter 7. Furthermore, analysis of RNA binding mutants addressed the requirement of a functional RRM for RBPMS localization to stress granules in HEK293 cells (Teplova et al., 2016). In parallel to RBPMS structure, another study resolved RBPMS2 structure underlying similar features between the paralogs (Sagnol et al., 2014).

Taken together the fact that RBPs can directly and indirectly affect splicing as well as being aware of the features underlying RBPMS binding to RNA, work described in this chapter set to unravel whether the splicing changes observed upon manipulation of RBPMS levels in PAC1 cells result from direct binding of RBPMS. To address this, bioinformatic and biochemical approaches were applied.

5.2 Results

5.2.1 RBPMS regulates ASEs enriched for tandem CAC motifs separated by variable spacer length

To address whether the ASEs identified in the PAC1 RNAseq were directly regulated by RBPMS, motif enrichment analyses were carried out by Dr. Aishwarya Jacobs using the Matt toolkit (Gohr and Irimia, 2018). Analyses were performed using

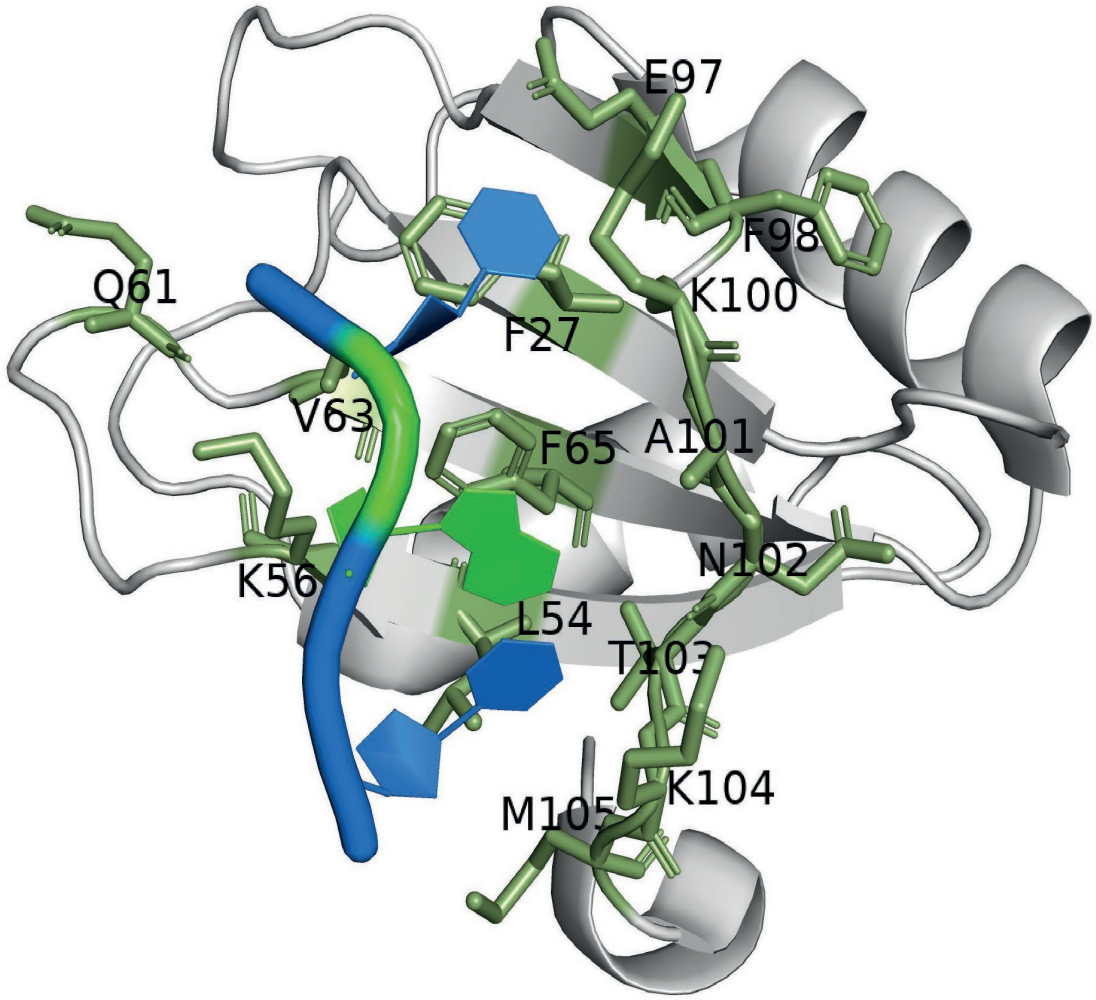


Fig. 5.1 Structure of RBPMS bound to CAC RNA molecule Crystal structure of RBPMS (gray) bound to a CAC RNA molecule (blue and green representing C and A nucleotides respectively). The residues involved in the RNA binding are labeled and highlighted in green. Image was generated using PyMol (v2.3.0) and the deposited structure of RBPMS RRM–RNA complex from the Research Collaboratory for Structural Bioinformatics PDB (accession code 5DET).

RBPMS binding motif (Farazi et al., 2014) and cassette exons regulated during PAC1 dedifferentiation, RBPMS knockdown and overexpression. Exonic and adjacent intronic regions were analyzed for enrichment revealing a distribution pattern similar to the position-dependent activity of other RBPs (Fu and Ares, 2014). Pairs of CACs separated by 1-12 nt ($CACN_{1\text{-}12}CAC$) were significantly enriched upstream (∼100 nt) and within exons repressed by RBPMS whereas exons activated showed enrichment in the downstream intron (Figure 5.2 and summarized in 5.3). The clearer position patterns in the RBPMS-map generated using the RBPMS-A overexpression dataset are explained by the total number of events found regulated in this experiment and consequently the number of events used in the motif enrichment analysis (2965 versus 99). Moreover, consistent with its role in AS during SMC switching, $CACN_{1\text{-}12}CAC$ enrichment was also identified around exons regulated during PAC1 dedifferentiation (Figure 5.2). Exons more included in the differentiated state displayed downstream enrichment and in those more included in the proliferative state, CAC motifs were found within and upstream of the regulated exon (Figure 5.2). Furthermore, significant depletion of RBPMS binding sites was also detected upstream and within exons repressed by RBPMS-A overexpression (Figure 5.2 and 5.3).

Therefore, RBPMS is likely to directly regulate AS of its targets since enrichment for its optimal recognition element was found around regulated exons. In addition, it is possible that binding to repressive areas is dominant over activation as indicated by the depletion of CACs in repressive locations for exons that are activated by RBPMS overexpression.

5.2.2 RBPMS regulation of splicing relies on the recognition of CAC motifs

To further validate the importance of RBPMS *cis*-elements in the regulated exons, HEK293 cells were co-transfected with minigene reporters and RBPMS expression vectors (Figure 5.4). Minigenes were representative of RBPMS-activated and RBPMS-repressed ASEs, *Flnb* and *Tpm1* respectively. Both regulated exons presented potential RBPMS motifs in the expected locations, downstream for activation and upstream for repression (Figure 5.5 and 3.11). Regulation of endogenous ASEs by RBPMS overexpression in HEK293 was previously established and described in Chapter 3. In addition to the fact that RBPMS was sufficient to switch to the differentiated SMC splicing pattern, RBPMS overexpression in HEK293 also highlighted some differential

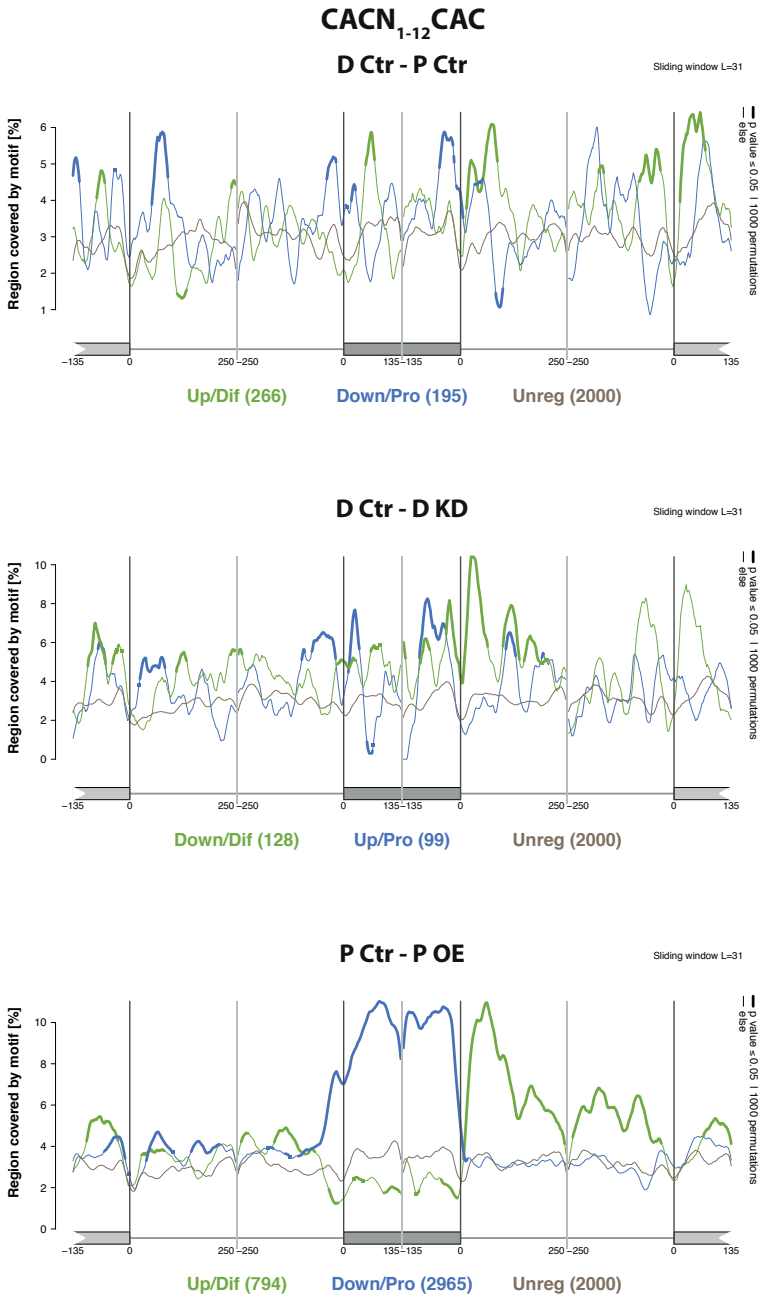


Fig. 5.2 RBPMS displays positional dependent splicing activity. RBPMS motif maps in regulated cassette exons from PAC1 dedifferentiation (D Ctr - P Ctr), **top**, RBPMS knockdown (D KD - D Ctr), **middle**, and RBPMS overexpression (P OE - P Ctr), **bottom**. A pair of CACs separated by 1 to 12 nucleotides was used as RBPMS binding site. Motif enrichment as well as depletion is shown for upregulated or differentiated exons (green), downregulated or proliferative exons (blue) and unregulated exons (gray). Motif maps were generated using the Matt toolkit. Statistical significance calculated by the Matt toolkit is shown by the line width.

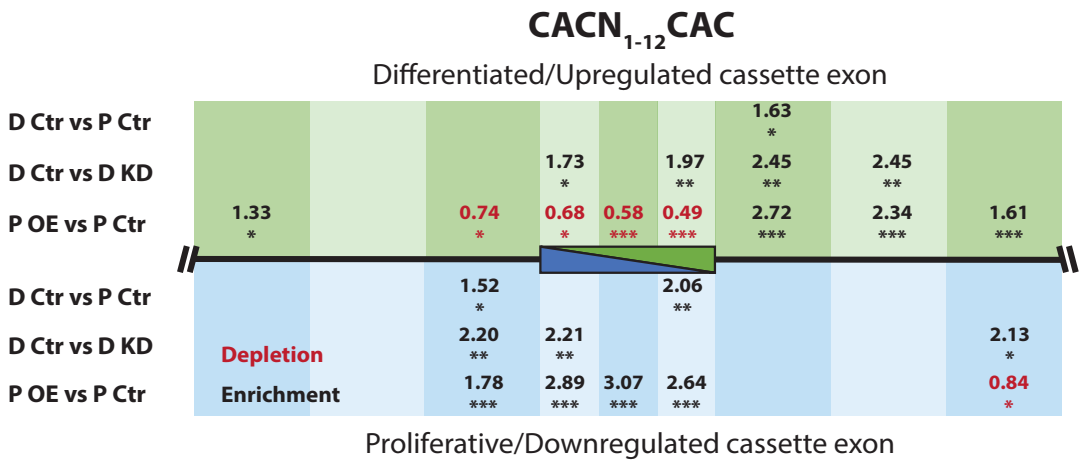


Fig. 5.3 RBPMS motif enrichment in PAC1 dedifferentiation and RBPMS experiments. Summary of RBPMS motif enrichment in the differentially spliced cassette exons from all the PAC1 experiments. PAC1 dedifferentiation (D Ctr - P Ctr), RBPMS knockdown (D Ctr - D KD), RBPMS overexpression (P OE - P Ctr). A pair of CACs separated by 1 to 12 nucleotides was chosen as RBPMS motif. Values in black represent fold enrichment and in red fold depletion. Top and bottom (green and blue respectively) of the panel represent upregulated and downregulated cassette exons. Statistical significance calculated by the Matt toolkit is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Note: RBPMS knockdown comparison was inverted to represent RBPMS repressed and activated events.

activity between isoforms. Therefore, regulation of the minigenes was tested upon both RBPMS-A and RBPMS-B isoforms and also its paralog, RBPMS2.

In a minigene context, *Flnb* exon H1 was included at a level similar to the endogenous regulation in HEK293 cells (∼7%, Figure 3.13), whereas RBPMS-A and RBPMS2 strongly promoted inclusion of exon H1 (greater that 74% inclusion), being sufficient to switch AS isoform to the long *Flnb* transcript (Figure 5.4). Despite showing significant changes upon RBPMS-B expression, activation was observed to a lower extent (∼half of the other proteins) (Figure 5.4). Yet RBPMS-B protein expression levels were higher than the other proteins (Figure 5.4 - western blots). Likewise, *Tpm1* minigene showed basal inclusion level close to 100% (Figure 5.4) comparable to the endogenous levels in HEK293 cells (Figure 3.13), in which exon 3 was nearly completely included. Once more, RBPMS-A was sufficient to cause a near complete switch to exon 3 exclusion (Figure 5.4). RBPMS 2 was the second more active protein causing skipping of 47% of exon 3 (Figure 5.4). Lastly, RBPMS-B was the least active protein, three fold less active than RBPMS-A (Figure 5.4), even though it was expressed at a similar level to RBPMS-A as shown by the western blots. Thus, both minigene reporters could reproduce the endogenous AS response to RBPMS-A overexpression, showing a similar isoform dependent activity, in which RBPMS-B has the lowest activity.

To uncover the *cis*-elements involved in the RBPMS-mediated splicing, CAC motifs downstream of *Flnb* and upstream of *Tpm1* alternative exons were mutated to CCC in the minigene reporters. This point mutation was shown to be sufficient to impair RBPMS binding (Farazi et al., 2014). The response to RBPMS was then assessed by overexpression of the most active isoform in HEK293 cells. Mutation of the individual clusters of CAC motifs downstream of exon H1 (12 motifs divided into three clusters - 5.6A and B) did not affect the basal splicing of *Flnb* exon H1 in HEK293 cells, yet it revealed the importance of the second and third clusters for activation by RBPMS-A (PSI$<$ 16%) (Figure 5.6C). Additionally, consistent with that, the minigene with all of the CACs mutated was completely unresponsive to RBPMS-A overexpression (Figure 5.6E). Fredderick Richards, a Part III student working under my day to day supervision in the laboratory, generated a *Tpm1* minigene reporter in which all nine CAC motifs upstream of *Tpm1* exon 3 were mutated to CCC (Figure 5.7A). CAC mutations in the *Tpm1* minigene reporter did not impair the basal exon inclusion observed in HEK293 cells, ∼100% inclusion of exon 3 (Figure 5.7A). However, mutation completely abrogated RBPMS repression of *Tpm1* exon 3 (PSI_{ex3}= 96.6). Moreover, intermediate effects were obtained with individual and combined clusters (Appendix

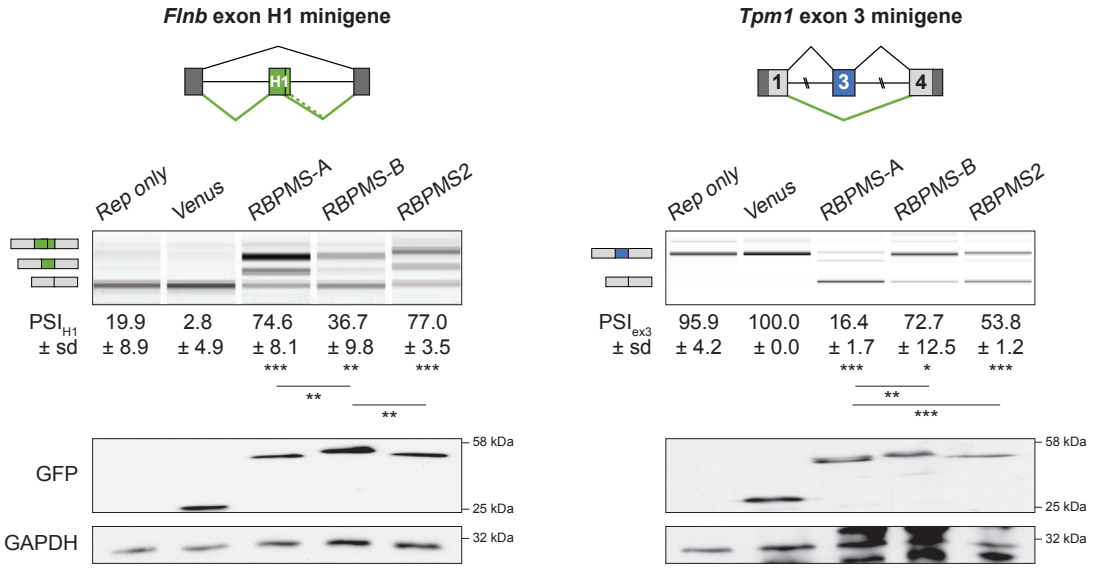


Fig. 5.4 RBPMS regulates splicing of *Flnb* and *Tpm1* minigenes. *Flnb* exon H1, left, and *Tpm1* exon 3 minigene reporters, right, co-transfected with RBPMS in HEK293 cells. Schematics of the minigenes are shown at the top. RT-PCRs for the splicing patterns. PSI values are shown as mean ± sd from n=3. For *Flnb*, both long and short isoforms were accounted for the final inclusion levels (PSI_{H1}). Isoform schematics on the side identify the PCR product. Differentiated exon is shown in green and proliferative exon in blue. Western blot against GFP to verify Venus tagged RBPMS levels. GAPDH was used as a loading control. Statistical significance from Student's t-test is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

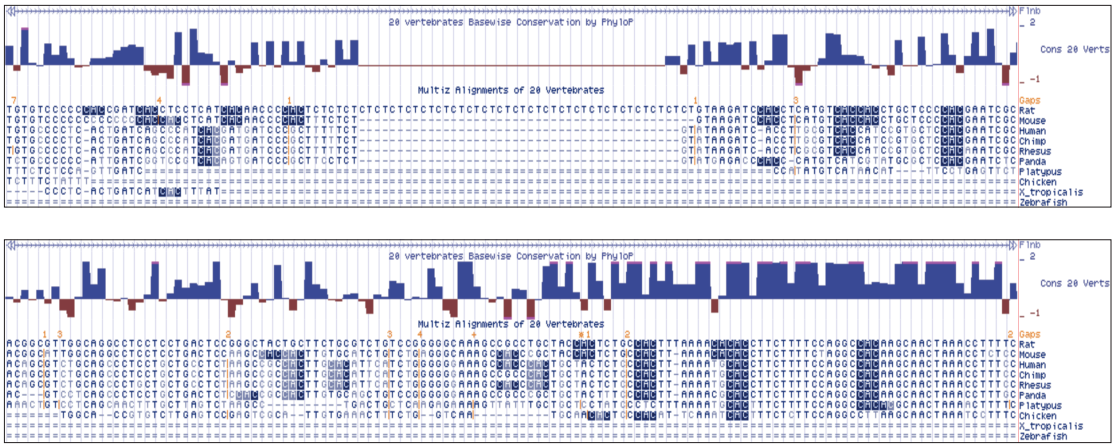


Fig. 5.5 Conserved CAC motifs downstream of *Flnb* exon H1. Conserved CAC motifs downstream of *Flnb* exon H1 are highlighted in blue. Intronic sequences shown are located 120 and 328 nt downstream of Flnb exon H1. Basewise conservation track and multiple alignment were generated by PhyloP and Multiz Alignments in the UCSC genome browser. Chr coordinates: chr15:18,780,016-18,780,147 and chr15:18,779,810-18,779,941.

- Figure A.3). Thus, CAC motifs adjacent to *Flnb* exon H1 and *Tpm1* exon 3 are required for activation and repression by RBPMS.

In the same manner, regulation of *Actn1* exons NM and SM were further investigated by transiently transfecting minigene reporters in HEK293 cells. Although rat *Actn1* minigene and endogenous *ACTN1* pairs of mutually exclusive exons were shown in Chapter 3 to be regulated by RBPMS in PAC1 and HEK293 cells respectively, the effects of overexpressed RBPMS could involve repression or activation of either the NM or SM exon respectively (Figure 3.12B). However, due to the conserved CAC motifs upstream of exon NM (Figure 3.11B and 5.8), it was hypothesized that RBPMS promoted the SM pattern by repressing exon NM. To test this, first HEK293 cells were co-transfected with the *Actn1* reporter containing both exons (Figure 5.9). As expected, exon NM was preferentially included in the HEK293 cells (PSI= 100%) and a complete switch towards the SM differentiated pattern promoted by RBPMS-A and RBPMS2 expression, but not RBPMS-B (Figure 5.9). Then, minigene reporters containing only one of the *Actn1* mutually exclusive exons (Southby et al., 1999) were tested upon RBPMS overexpression (Figure 5.9).

The *Actn1* reporter containing only the NM exon reproduced the complete switch upon RBPMS-A and 2 expression (PSI= 0.0%) but not B (PSI= 73.8%) (Figure 5.9). This suggests that RBPMS acts on *Actn1* AS by inhibiting inclusion of the NM

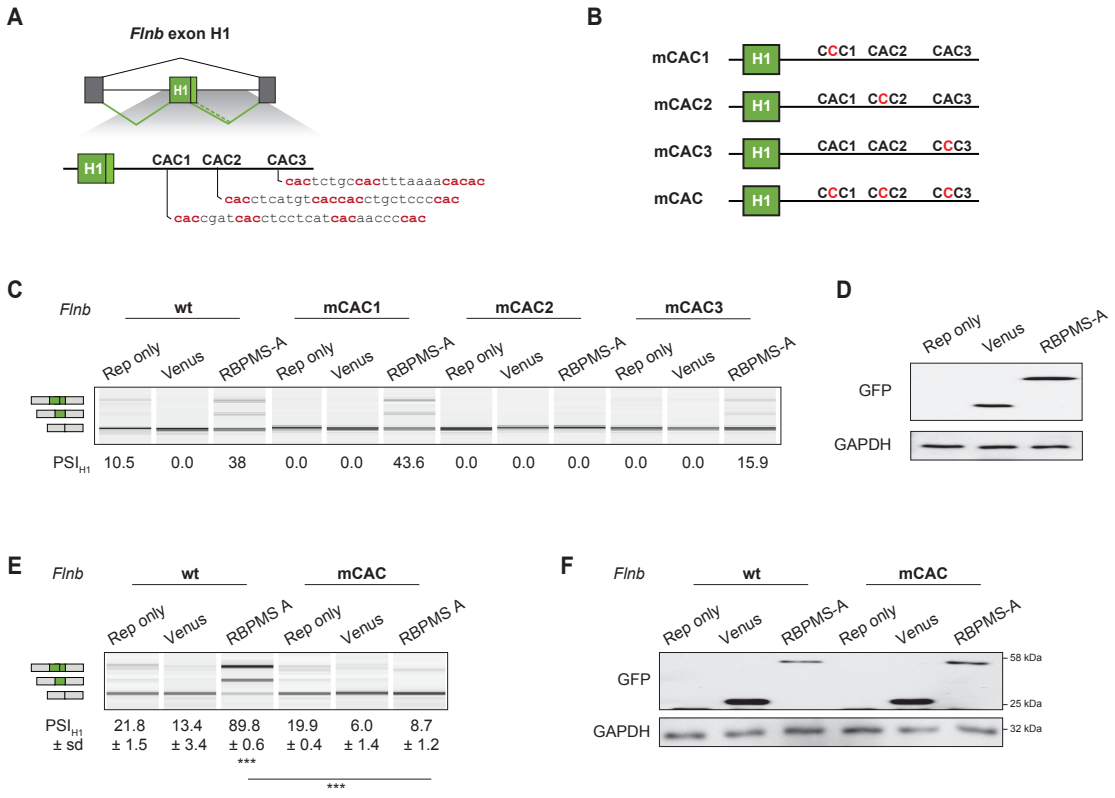


Fig. 5.6 CAC motifs downstream of *Flnb* exon H1 are required for RBPMS activation. A Schematic of the rat *Flnb* exon H1 minigene reporter. Potential RBPMS binding sites downstream of exon H1 are highlighted. **B** Schematic of the *Flnb* mCAC mutant minigenes (CAC to CCC). **C** RT-PCRs from HEK293 transient transfection with RBPMS-A and the different *Flnb* single mCAC mutants. Experiment from n=1. **D** Western blot for GFP to verify overexpression of Venus tagged RBPMS-A. **E** RT-PCRs from HEK293 transient transfection with RBPMS-A and the *Flnb* minigene reporter in which all the CACs were mutated. Values shown correspond to PSI ± sd (n= 3). **F** Western blot for GFP to verify overexpression of Venus tagged RBPMS-A. In **C** and **F**, reporter only and Venus controls were analyzed in parallel. *Flnb* isoforms schematics on the left indicate the PCR products. Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. In **D** and **F**, GAPDH was used a loading control.

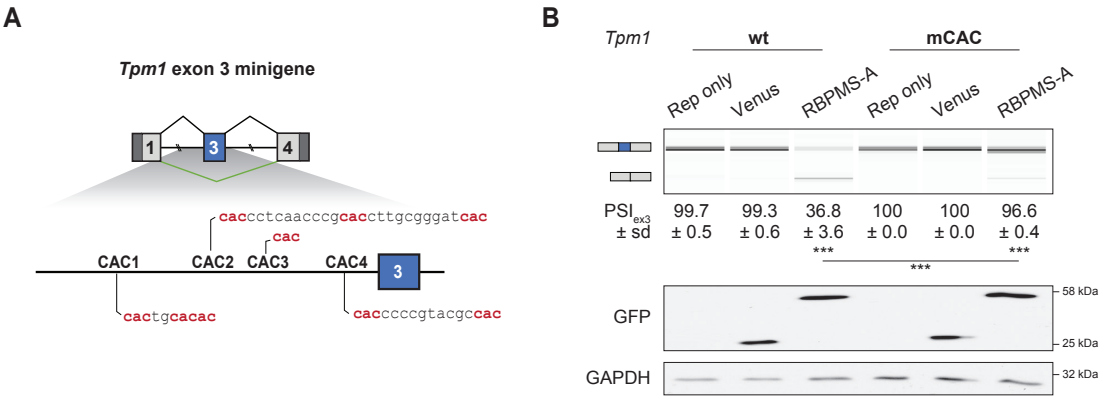


Fig. 5.7 CAC motifs upstream of *Tpm1* exon 3 are required for RBPMS repression. A Schematic of *Tpm1* exon 3 minigene with upstream CAC clusters highlighted. **B** RT-PCRs from HEK293 transiently transfected with RBPMS-A and the *Tpm1* minigene reporter wild-type (wt) and mCAC in which all the CACs were mutated. Values shown correspond to PSI ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Bottom, western blot against GFP to verify overexpression of Venus tagged RBPMS-A. GAPDH was used as a loading control. Protein size markers are indicated on the right.

exon. Nevertheless, the minigene containing only exon SM only produced the SM included transcript in HEK293 cells despite the absence of an effector. In that way the possibility that RBPMS also directly promotes inclusion of the exon SM could not be fully excluded. Thus, these data further confirm the requirement of nearby CACs for repression of the *Actn1* NM exon by RBPMS and further mutations of these sites will shed light on the importance of each clusters.

Thus, by repressing the NM exons of the two mutually exclusive events, *Tpm1* and *Actn1*, RBPMS promotes the SMC pattern in these two events. Moreover, RBPMS is also able to switch *Flnb* splicing to the SMC pattern by activation of the exon H1. In all of these events regulation by RBPMS was mediated by adjacent CAC motifs. Finally, consistent with previous experiments carried out in this study, RBPMS-B was the least active protein.

5.2.3 RBPMS directly binds to *Flnb* and *Tpm1* RNAs via CAC motifs

To confirm that the previously mutated CACs in the minigene reporters were actual RBPMS binding sites, *in vitro* binding assays were performed using *in vitro* transcribed

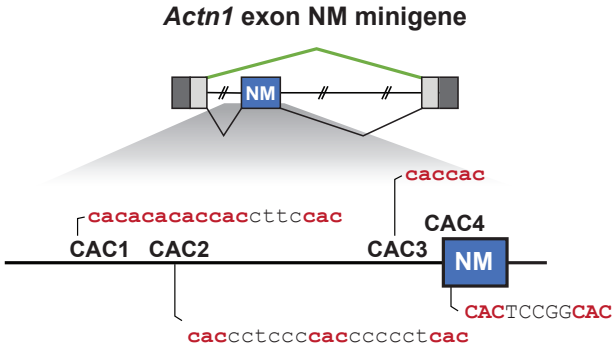


Fig. 5.8 *Actn1* contains CAC motifs within and upstream of exon NM. Schematic of the rat *Actn1* minigene reporter containing only exon NM with CAC clusters highlighted.

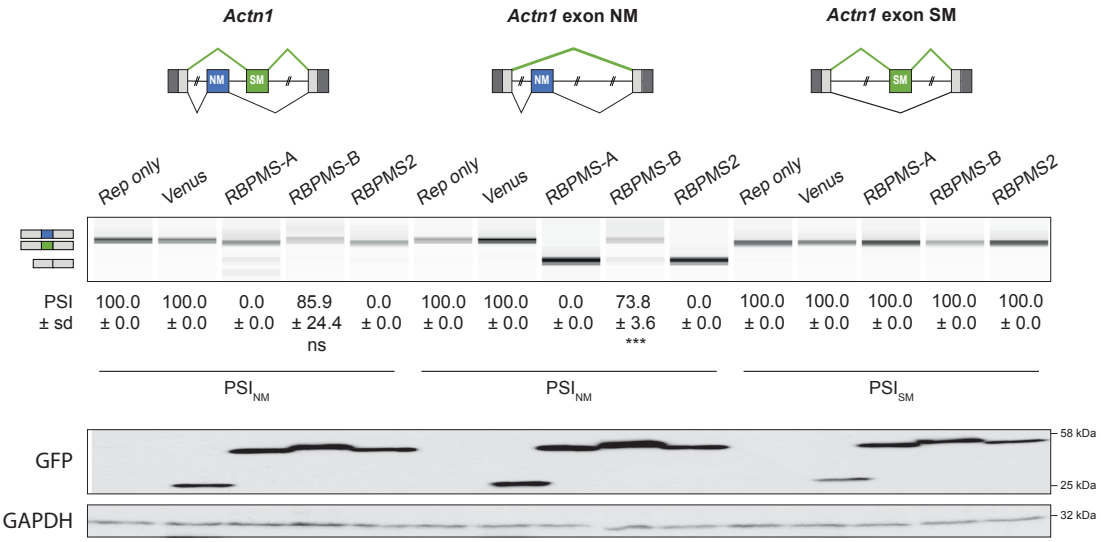


Fig. 5.9 RBPMS represses the NM exon of *Actn1*. HEK293 cell co-transfected with RBPMS isoforms and rat *Actn1* minigenes containing the pair of mutually exclusive exons (*Actn1*) or either one of them (*Actn1* exon NM or SM). Schematics of the minigenes are found at the top. RT-PCR products are identified by schematics on the left. PSI values shown are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Bottom, western blot against GFP to verify overexpression of Venus tagged RBPMS. GAPDH was used as a loading control. Protein size markers are indicated on the right.

RNAs and recombinant RBPMS-A and RBPMS-B proteins. Recombinant proteins were obtained as detailed in Chapter 2. Briefly, recombinant proteins containing a T7 and a 3xFLAG N-terminal tags and a His_6 tail were expressed in *E. coli*, followed by purification through two column systems. Isoform B was purified by a single histidine-affinity chromatography step. Recombinant proteins were visualized in a polyacrylamide gel and sequence verified by western blot (Figure 5.10). Additionally, MALDI-TOF mass mapping analysis of both recombinant proteins, further validated their identity and also eliminated any post-translational modification (Figure 5.10).

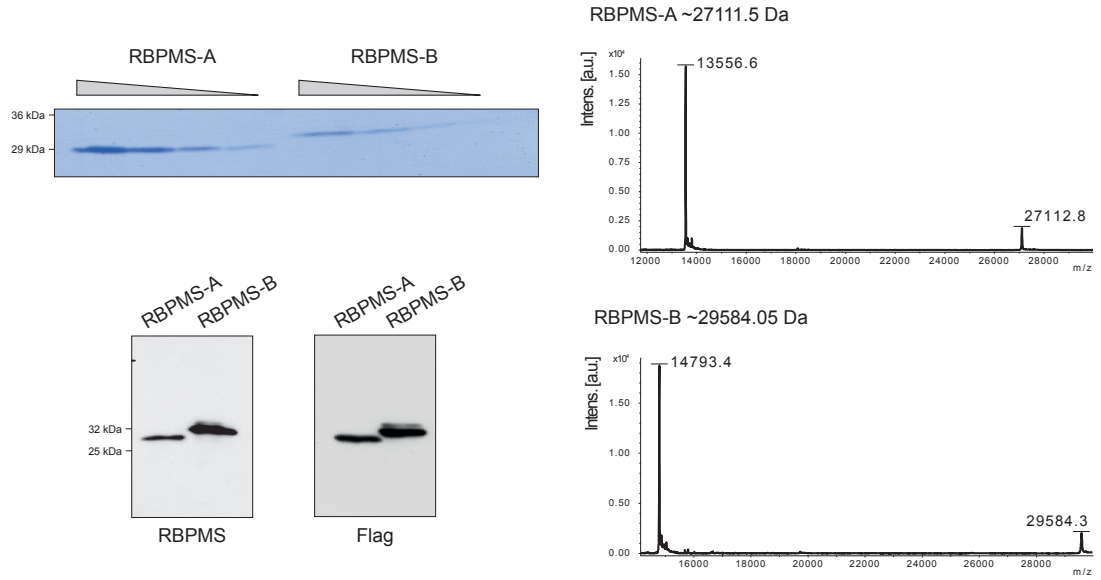


Fig. 5.10 Rat RBPMS-A and RBPMS-B recombinant proteins. Left, quality of RBPMS recombinant proteins were verified in polyacrylamide gel stained with coomassie. Recombinant protein was loaded into the gel in a 1:2 serial dilution (2 - 0.25 μM). Specificity of the amino acid sequence of recombinant RBPMS proteins was also verified by western blot using RBPMS and FLAG antibodies. Size markers are indicated on the left. **Right**, RBPMS recombinant protein was verified by MALDI mass spectrometry analysis and is shown for RBPMS-A and RBPMS-B. The two peaks represent double and single charged RBPMS.

RBPMS binding to rat *Flnb* and *Tpm1* CAC clusters was verified by using both electrophoretic mobility shift assay (EMSA) and UV-crosslinking (Figure 5.11 and 5.12). RBPMS-A and B isoforms bound to *Flnb* with similar apparent affinities (Kd $\sim$0.5 μM) (Figure 5.11). On the other hand, no binding was observed to the CAC mutant RNAs (*Flnb* mCAC) (Figure 5.11). Moreover, binding to the wt but not to the mutant *Flnb* RNA was also reproduced by UV-crosslinking, although in this assay, RBPMS-A

showed greater intensity over RBPMS-B (Figure 5.11). For *Tpm1*, both RBPMS isoforms bound to the wt RNA similarly to the binding observed for *Flnb* (Figure 5.12). Although CAC mutations reduced the formation of the bound complex, they were not sufficient to completely abrogate RBPMS binding (Figure 5.12). UV-crosslinking for *Tpm1* RNAs was very inefficient, although a faint band was observed for the wt but not for the mutant RNAs (Figure 5.12). These data corroborate the fact that RBPMS regulates splicing by directly binding to CAC elements downstream of activated exons (*Flnb*) and upstream of repressed exons (*Tpm1*).

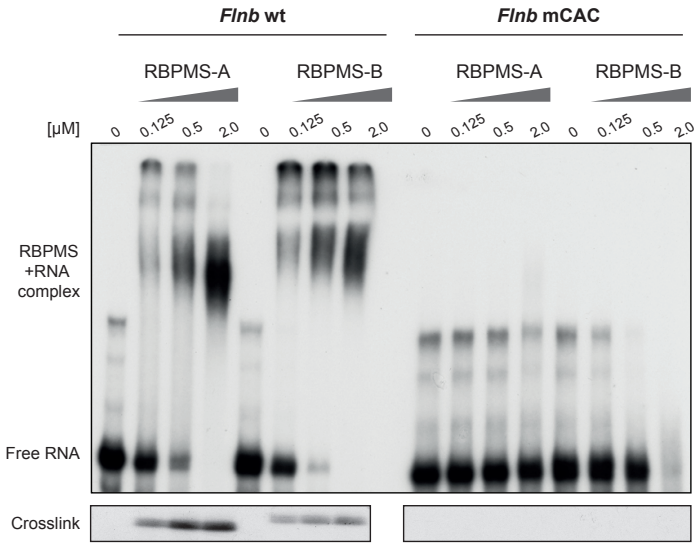


Fig. 5.11 RBPMS binds to *Flnb* RNAs via CAC motifs. EMSA (**top**) and UV-crosslink (**bottom**) using recombinant RBPMS-A and RBPMS-B and *in vitro* transcribed *Flnb* RNA, wild-type and mCAC, corresponding to the intronic region containing CAC clusters downstream of exon H1. *in vitro* binding experiments were performed using recombinant protein in a serial dilution (1:4; 2 to 0.125 μM).

5.2.4 RBPMS binding mutant is not able to regulate splicing

Finally, to further confirm the direct regulation of splicing by RBPMS, an RNA binding mutant version of RBPMS-A isoform was transiently transfected in HEK293 cells and splicing changes assessed (Figure 5.13). The mutation of the lysine in the position 100 to a glutamic acid was previously reported to disrupt RBPMS binding to RNA and therefore it was chosen for this study (Farazi et al., 2014). Consistent with all the data presented so far, RBPMS wt was able to repress the NM exon from the mutually

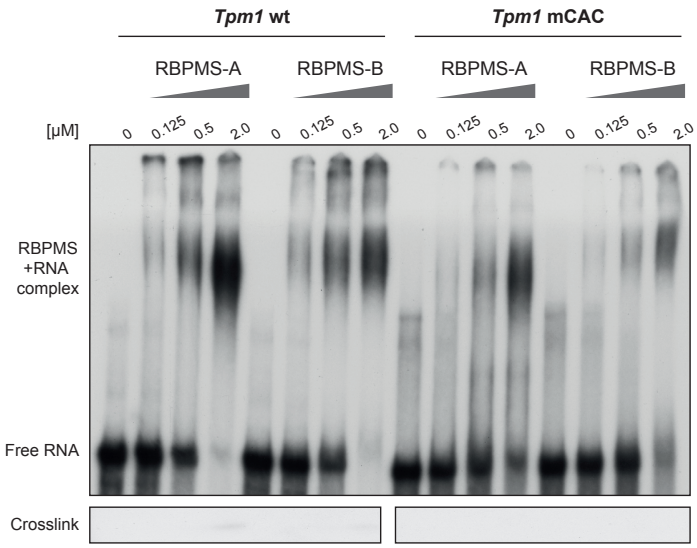


Fig. 5.12 RBPMS bind to *Tpm1* RNAs via CAC motifs. EMSA (**top**) and UV-crosslink (**bottom**) using recombinant RBPMS-A and B and *in vitro* transcribed *Tpm1* RNA, wild-type and mCAC, corresponding to the intronic region containing CAC clusters upstream of exon 3. *in vitro* binding experiments were performed using recombinant protein in a serial dilution (1:4; 2 to 0.125 μM).

exclusive events in the rat *Tpm1* minigene reporter and endogenous *ACTN1* as well as activate splicing of endogenous *FLNB* and *MPRIP* cassette exons (Figure 5.13). In contrast, the RNA binding mutant, which was expressed as similar protein levels than the wt RBPMS-A, showed no splicing activity upon any of the four ASEs tested (Figure 5.13). Once again these data show that RBPMS regulation of splicing, either activation or repression, relies on direct binding to its RNA targets, with regulation being lost upon disruption of the protein-RNA interaction.

Therefore, RBPMS is able to regulate splicing of its pre-mRNA targets by directly binding to CAC sites present upstream or downstream of repressed and activated exons, as shown by the impaired splicing upon disruption of either the *cis*-elements in the minigene reporters or of the RNA binding capacity of RBPMS.

5.3 Discussion

RBPMS: a direct regulator of alternative splicing

RBPMS *in vitro* binding assays revealed that both isoforms could bind to *Flnb* and *Tpm1* RNAs and mutation of CAC motifs directly impacted binding of RBPMS (Figure

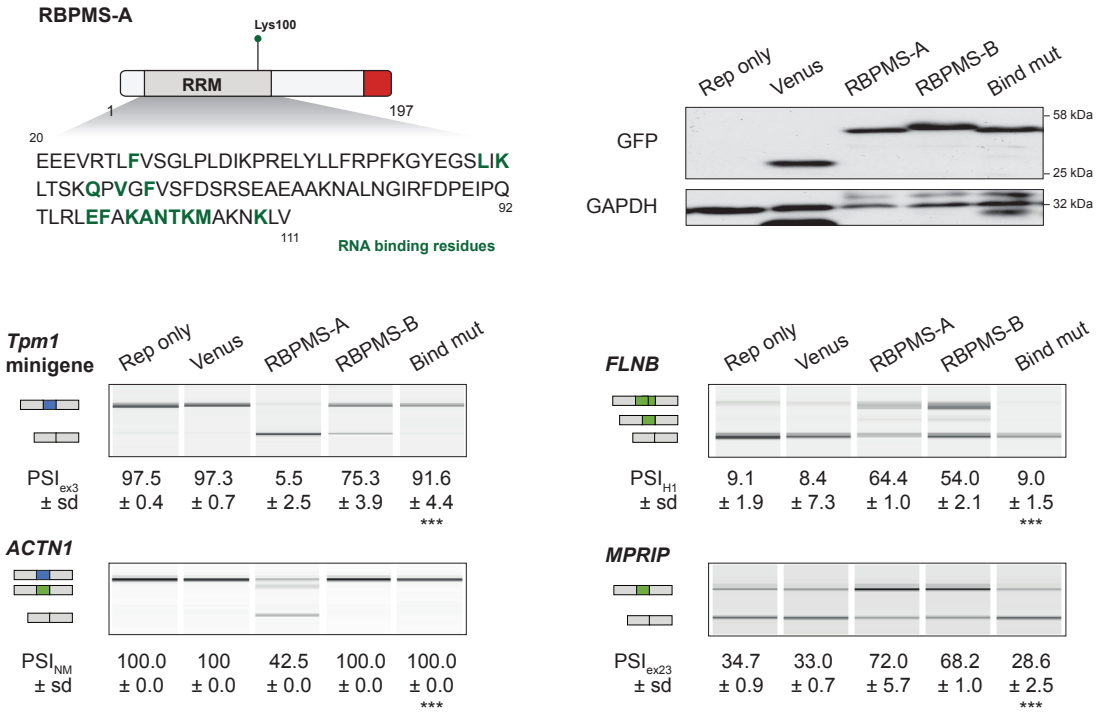


Fig. 5.13 RBPMS binding mutant is not able to regulate splicing. RBPMS-A binding mutant (K100E from Farazi et al. (2014)) was tested for regulation of rat *Tpm1* minigene reporter and endogenous *ACTN1*, *FLNB* and *MPRIP* splicing in HEK293 cells. Wild-type RBPMS-A and RBPMS-B were tested alongside as well as the reporter only and Venus controls. Schematic of RBPMS-A protein domain arrangement, RRM amino acid sequence is found below it. Residues involved in RNA binding are highlighted in green. RT-PCR products are identified by schematics on the left. PSI values shown are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Western blot against GFP to verify overexpression of Venus tagged RBPMS. GAPDH was used as a loading control. Protein size markers are indicated on the right.

5.11 and 5.12). The requirement of RNA binding was further confirmed *in vivo* by the overexpression of an RNA binding mutant RBPMS-A, which was not able to promote splicing changes as the wild-type (Figure 5.13). Consistent with that, a functional RRM has previously been shown to be necessary for RBPMS localization to stress granules in HEK293 cells (Farazi et al., 2014). Other studies have also reported similar results (Gerber et al., 2002; Hörnberg et al., 2013; Kaufman et al., 2018), however because the RBPMS mutant used had the full RRM deleted, it is not possible to untangle whether the loss of function resulted from impairing RBPMS RNA binding or dimerisation. Thus, the data presented here support RBPMS as a direct regulator of AS in SMC.

The findings, however, are contrary to the results from Farazi et al. (2014), in which no major role for RBPMS in mRNA splicing was found after RBPMS manipulation in HEK293 cells and following transcriptional profiling. These opposite results are likely to arise from the combination of the inefficient UV crosslinking *in vitro* and elimination of ASE that although detected in the RNAseq did not display CLIP tags. *In vitro* crosslinking experiments carried out for RBPMS and RNA targets (*Flnb* and *Tpm1*) were very inefficient, yet band shift experiments showed formation of RBPMS-RNA complexes (Figure 5.11 and 5.12). Therefore, targets like *Actn1* and *Flnb* that are regulated in HEK293 cells might not have presented PAR-CLIP tags in Farazi et al. (2014) due to the limitations of the crosslinking approach. Although the PAR-CLIP was carried out with 4-thioU which is supposed to be more efficient. Nevertheless, the use of minigene constructs for these splicing events (*Tpm1*, *Actn1* and *Flnb*) proved functional binding of RBPMS. For the future, CLIP should be applied to investigate global RNA targets of RBPMS in more relevant biological system. For instance, PAC1 cells or embryonic stem cell derived VSMCs (ES-VSMCs) (Cheung et al., 2014) in which more target transcripts would be expressed at high levels.

Interestingly, RBPMS-B lower activity in splicing could not be explained by impaired binding as indicated by the EMSAs (Figure 5.11). Therefore, it is possible that RBPMS repressive function is mediated by RBPMS-A C-terminus residues (20 last amino acids) via specific protein-protein interaction. Further structural and functional analysis attempting to resolve the differential activity are described in Chapter 7.

On the other hand, the direct binding of RBPMS to a few of its AS targets does not completely eliminate the possibility of indirect ASEs among the events identified regulated upon manipulation of RBPMS levels. In fact, it is likely that RBPMS regulates other splicing factors that could then trigger a cascade of secondary AS changes, as suggested for a master splicing regulator (Jangi and Sharp, 2014). Thus,

this possibility is addressed in the following chapter.

Tpm1 alternative splicing regulation by RBPMS

RBPMS repressed inclusion of *Tpm1* exon 3 by using upstream CAC motifs. Despite using a minigene reporter containing exon 3 only, this ASE in the *Tpm1* gene consists of a pair of mutually exclusive exons, exon 2 and 3. Exon 2 is selected in SMCs and exon 3 in skeletal muscle, heart and brain (Wieczorek et al., 1988). Consequently, splicing of *Tpm1* exon 2 and 3 has been used as a differentiation marker in SMC AS studies (Gooding and Smith, 2008; Llorian et al., 2016; Rothman et al., 1992). In spite of that, the functional consequences of this ASE to the protein and cell biology remain unknown. Moreover, regulation of *Tpm1* exon 3 was previously studied in the laboratory and showed to be under control of MBNL1 and PTB (Gooding et al., 2013). On the other hand, neither of these proteins, MBNL1 and PTB, were sufficient to cause the switch in mRNA isoforms observed in RBPMS overexpression in HEK293. Thus, it is likely that these proteins co-operate to establish the SMC pattern of *Tpm1* splicing.

Flnb alternative splicing regulation by RBPMS

RBPMS promoted the activation of the *Flnb* exon H1 via binding to downstream CAC motifs. *Flnb* is an actin binding and adhesion protein and its exon H1 encodes the H1 hinge domain (van der Flier et al., 2002). Moreover, splice variants of FLNB differentially modulate the organization of actin cytoskeleton and binding to integrins with direct effects on myogenesis of C2C12 cells (van der Flier et al., 2002). More recently, the same exon of *Flnb* was described under regulation of another splicing RBP, QKI, during epithelial-mesenchymal transition (EMT) (Li et al., 2018). RBPMS and QKI antagonistic control of *Flnb* exon H1 could potentially be a common feature of the regulation of the SMC AS program, since QKI has been observed to promote the proliferative phenotype (Llorian et al., 2016; van der Veer et al., 2013).This potential interplay between RBPs was further explored in the following chapter.

Actn1 alternative splicing regulation by RBPMS

Regulation of *Actn1* splicing was previously investigated in this laboratory. It contains a pair of mutually exclusive exons, which are tissue specifically selected: a non-smooth muscle, NM, and a smooth muscle, SM exon (Waites et al., 1992). Additionally, the skipping of both exons is also observed in small proportions in SMCs (Gromak et al., 2003). The NM exon encodes a functional Ca^{2+}-binding EF hand domain which

is then more present in non-SMCs. On the other hand, the ACTN1-differentiated isoform lacks this full domain due to the inclusion of exon SM instead. The impairment of Ca^{2+}-binding allows stabilization of the ACTN1 containing structures in contractile cells (Waites et al., 1992). Overexpression of RBPMS in HEK293, particularly isoform A, promoted a complete switch towards the exon NM skipping in the minigene reporters containing both exons and exon NM only, consistent with an RBPMS repressor activity upon exon NM. Although previously described to be regulated by PTBP1 and CELF proteins, none of these proteins were able to produce such a strong effect on *Actn1* splicing (Gromak et al., 2003; Southby et al., 1999). Interestingly, synergistic and antagonistic interactions between CELF and PTB proteins regulated *Actn1* splicing, in which CELF protein family members activated SM exon and also repressed exon NM by displacement of PTBP1, a repressor of exon SM (Gromak et al., 2003).

Therefore, taken the fact that RBPMS and other RBPs (e.g. MBNL1, PTB and QKI) regulates common targets, further investigations focusing on the synergistic and antagonistic interactions between families of regulators could shed some light on the control of the differentiated SMC AS program.

5.4 Final conclusions

This chapter aimed to determine whether RBPMS could directly affect splicing. The strong enrichment of tandem CAC motifs with RBPMS regulated exons (Figure 5.2 and 5.3), the functional binding of RBPMS to *Flnb* and *Tpm1* (Figure 5.11), the distinct positional signatures of splicing activation and repression mediated by RBPMS (Figure 5.2 and 5.3) and the loss of splicing function by the RNA binding mutant (Figure 5.13) support RBPMS as a direct regulator of splicing in differentiated SMC.

In conclusion, the main findings of this chapter are:

1. RBPMS-regulated exons display strong enrichment of tandem CACs motifs.
2. RBPMS-map revealed distinct positional signatures for splicing activation and repression by RBPMS.
3. EMSA and UV-crosslinking confirmed binding of RBPMS to *Flnb* and *Tpm1* RNAs.
4. RBPMS represses splicing via binding to upstream CAC motifs whereas activation requires downstream motifs.

5. CAC motifs show compensatory effects demanding mutation of several sites for disruption of RBPMS regulation.

6. RBPMS represses the NM exon from the pair of mutually exclusive exons in the *Actn1* transcripts.

7. RBPMS RNA binding mutant is not able to regulate splicing.

8. RBPMS isoforms exhibit differential activity; RBPMS-B less active.

In addition, this chapter also led to other questions listed below:

- **What is the specific mechanism behind RBPMS regulation of splicing?**
- **How can the isoforms display different splicing activity if both RBPMS-A and RBPMS-B can bind RNA targets *in vitro*?**
- **What are the synergistic and antagonistic interactions between RBPMS and other RBPs in the regulation of common pre-mRNA targets (e.g. *Tpm1*, *Actn1* and *Flnb*)**?

The next chapters in this study provide some insights into these questions.

Chapter 6

RBPMS: a regulator of regulators

6.1 Introduction

6.1.1 Contribution of splicing to establishment of robust transcriptomes

Studies of the regulation of transcriptomes and their role in cell homeostasis and differentiation have focused on transcriptional programs as the main player in the establishment of tissue transcriptomes. However, RNA sequencing technologies have advanced in the past decades and provided evidence for the contribution of post-transcriptional processes, especially AS, in the regulation of tissue transcriptomes. AS is able to regulate a network of highly interconnected events, providing stability to the system without losing its ability to respond to external stimuli (Jangi and Sharp, 2014).

AS reshapes the cell transcriptome by generation of alternative protein isoforms and down-regulation of transcripts via coordination of AS-NMD (nonsense-mediated mRNA decay). The combination of these processes can then form different regulatory units such as negative and positive feedback loops (Figure 6.1). Furthermore, cross-regulation (Figure 6.2) as well as integration of these individual regulatory motifs into splicing cascades (Figure 6.2) provide mechanisms required for building robust networks in biological systems (Jangi and Sharp, 2014). The different regulatory units of the splicing network are further discussed below.

6.1.1.1 Negative feedback of RBPs

AS coupled with NMD regulates the steady-state of many genes including many RBPs. These RBPs can then be engaged in negative autoregulation through regulation of the splicing of exons containing in-frame premature stop codons within their own transcript. Thus RBPs reduce their own expression levels by targeting their own transcripts to NMD (Figure 6.1). In fact, many splicing factors show enrichment for conserved AS-NMD events, revealing a common mechanism of gene expression control of RBPs (Jangi and Sharp, 2014; Wollerton et al., 2004). RBPs that display negative autoregulation comprise for example the splicing regulators SR and hNRNP proteins (Lareau et al., 2007; Ni et al., 2007). SR proteins usually promote poison exon inclusion whereas some hnRNPs promote frameshift exon skipping e.g. PTBP1.

6.1.1.2 Positive feedback of RBPs

In contrast to negative feedback, RBP direct positive autoregulation via productive splicing is very unusual. Interaction between two distinct splicing factors could provide the mechanisms for positive regulation (Figure 6.1). Double-negative feedback loops combine two RBPs that negatively cross-regulate each other, generating mutually exclusive expression of these factors. Positive feedback generally produces bistability in which two distinct steady states can be achieved (Jangi and Sharp, 2014). Nevertheless, to date, AS-NMD positive feedback loop like described above has mainly been reported to operate in insects (see section 6.1.2) (Pervouchine et al., 2018; Salz, 2011).

Another mode for establishing positive feedback is by coupling splicing and transcription factors (Jangi and Sharp, 2014). For instance, during development, OCT4, SRSF2 and MBD2 form a composite positive feedback loop, in which OCT4 drives expression of the splicing factor SRSF2 which affects splicing of MBD2 towards an isoform that promotes OCT4 expression (Lu et al., 2014).

6.1.1.3 RBP cross-regulation

Steady-states of RBPs under autoregulatory negative feedback are susceptible to additional post-transcriptional regulatory inputs (Jangi and Sharp, 2014). This secondary modulation is responsible for defining the final expression level of the RBP across different cell types. One mechanism involved in this process is cross-regulation between RBPs, such as inhibition of an RBP negative autoregulatory poison exon upon binding of a second RBP (Figure 6.2) (Jangi and Sharp, 2014). In this manner, a more robust and tunable expression pattern of RBPs is ensured.

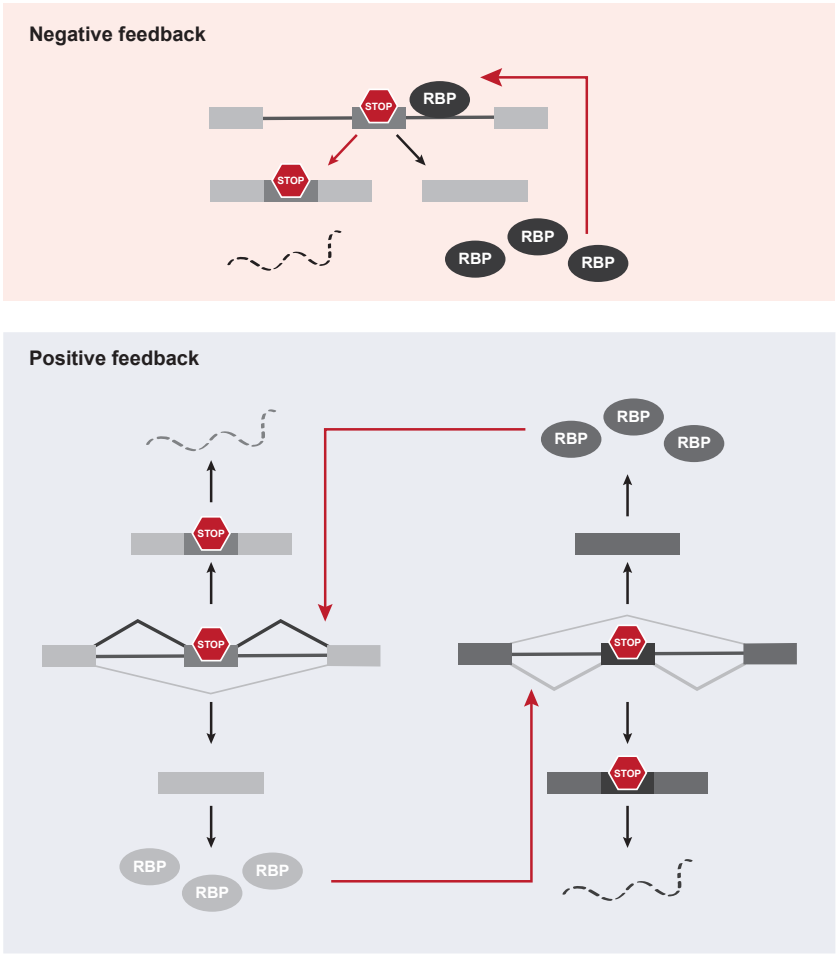


Fig. 6.1 Regulatory units in splicing networks. Top, autoregulatory negative feedback of RBPs mediated by regulation of NMD targeted exons. Regulation of poison exons (red hexagon indicates premature termination codon, PTC) within its own transcript leads to reduction of its own protein levels, maintaining a steady-state level of expression, e.g. SRSF genes. Nonetheless, exon skipping can also lead to frame-shifting and subsequent downregulation of transcripts e.g. PTBP1 (Wollerton et al., 2004). **Bottom**, positive feedback via double-negative feedback loop. This results from mutually exclusive expression of two RBPs due to their negative cross-regulation.

Thereby it is expected that cross-regulation between RBPs leads to secondary splicing changes within the splicing network of an specific RBP (Figure 6.3). Although the network of a splicing factor primarily consists of its direct targets, indirect events are also identified due to regulation of secondary RBPs levels through AS-NMD. Thus, the initial signal from the first RBP can be amplified by the triggering of a cascade of indirect splicing events (Jangi and Sharp, 2014). Such regulation was observed during neuronal differentiation in which RBFOX2 upregulates PTBP2 by hindering the autoregulatory AS-NMD suppression of *Ptbp2* (Jangi and Sharp, 2014).

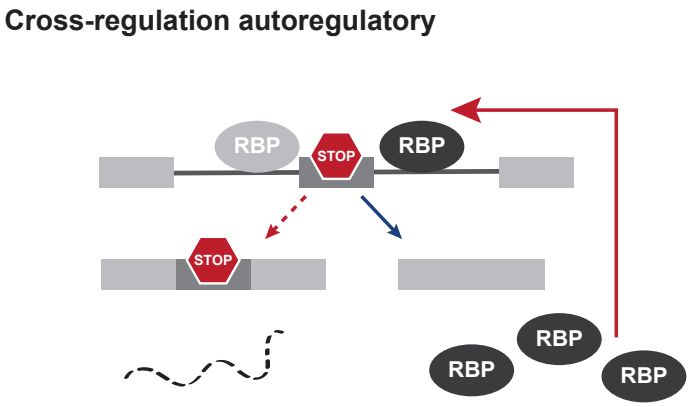


Fig. 6.2 RBP autoregulatory feedback cross-regulation. Autoregulatory negative feedback loops can be controlled by other RBPs. Thus, a higher expression level of the auto-regulated RBP can be achieved by a secondary RBP that inhibits the negative regulation.

Therefore a better understanding of these regulatory networks is critical to gain more insights into how splicing changes can determine cell fate.

6.1.2 Master regulators regulate other splicing regulators

RBPs acting as master regulators of splicing are proposed to be positioned at the top of a splicing cascade, triggering direct changes and also indirect events by regulation of secondary splicing factors (Jangi and Sharp, 2014). In this manner, master splicing regulators would be able to build a robust network with a critical role in cell fate determination.

Despite the fact that master RBPs are likely to be downstream of larger transcriptional programs, as indicated by the approach of using super-enhancers for their

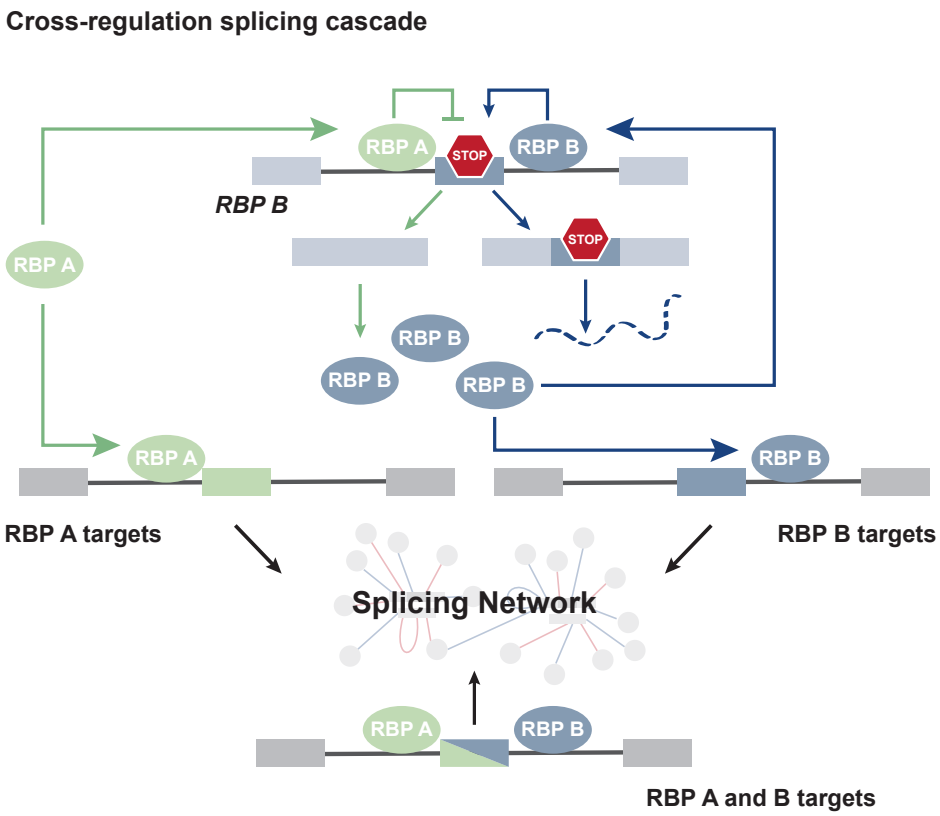


Fig. 6.3 Cross-regulation establishes robust splicing networks by splicing cascades. RBPs are able to amplify their splicing networks by combining direct regulated targets to cross-regulated events that trigger a cascade of secondary events. Thus, this integrated system amplifies the original signal and fine-tunes the RBP splicing network.

identification, transcription factors can also be found under the control of splicing factors (Jangi and Sharp, 2014). A very well characterized interplay of transcriptional and post-transcriptional factors is the sex determination in *Drosophila* (Salz, 2011). In the fruit-fly, the splicing repressor Sex Lethal (Sxl) is a key RBP in the sex-determination pathway by activating the female isoform of the RBP Transformer (Tra) and repressing the male-specific lethal 2 (Msl2) transcript. Moreover, female-specific isoforms of the transcription factors Dsx and Fru are produced by the female-Tra isoform (Salz, 2011). Thus, Sxl is a master splicing regulator that is required and even sufficient in the determination of cell identity in *Drosophila*.

Therefore, master regulators can activate a robust splicing network by integrating direct and indirect splicing changes in cooperation with alteration of gene expression which arises from AS-NMD as well as regulation of transcription factors (Jangi and Sharp, 2014). Understanding this co-regulation between master regulators and other RBPs is critical for uncovering their roles in determining tissue specific splicing programs. Secondly, the contribution of other RBP functions, e.g. mRNA stability and localization, also need to be addressed to resolve the additional role of various post-transcriptional regulatory mechanisms in the establishing of cell-specific transcriptomes.

Thus, taken the fact that master regulators can indirectly affect splicing and gene expression by modulation of other RBPs and transcription factors (Jangi and Sharp, 2014), this chapter set to identify and characterize any other splicing factor under RBPMS regulation. Moreover, we also investigated if RBPMS targets other post-transcriptional and transcriptional factors.

6.2 Results

6.2.1 Regulators regulated by RBPMS

To identify the protein classes targeted by RBPMS, functional classification of genes differentially spliced upon manipulation of RBPMS levels in PAC1 cells was carried out using PANTHER (v14.1) (Figure 6.4 and 6.5). In addition to the enriched functions previously discussed (cytoskeletal, cell adhesion and ECM proteins), both datasets showed regulation of transcription factors and nucleic acid binding proteins, including RBPs (Figure 6.4 and 6.5). RBPMS knockdown affected seven RBPs, nine DNA binding proteins and twelve transcription factors (Figure 6.4) whereas those numbers were much larger in the overexpression experiment 91, 46 and 118 respectively (Figure 6.5).

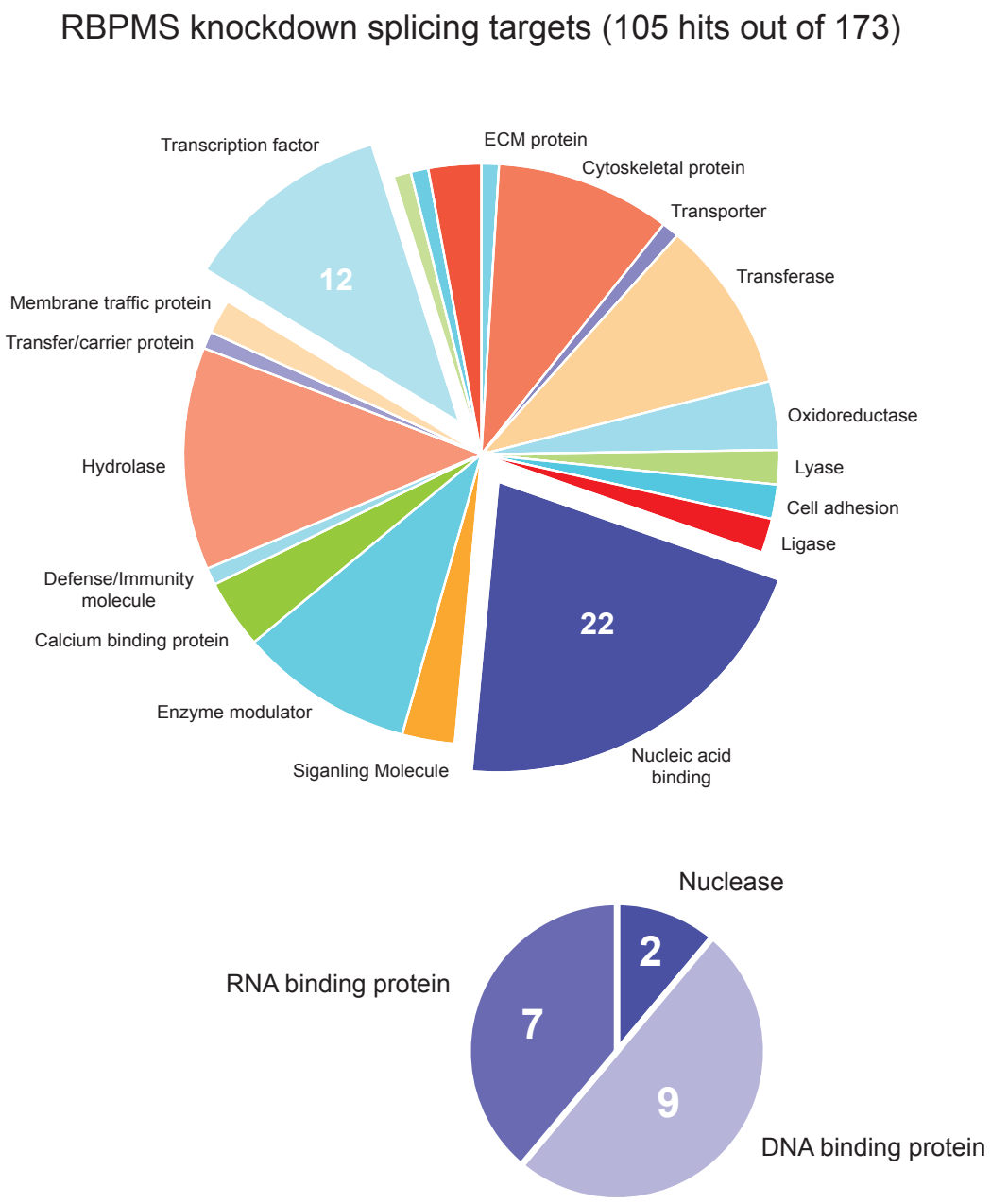


Fig. 6.4 RBPMS knockdown affects splicing of genes coding for a range of protein classes. Functional classification of RBPMS knockdown splicing targets according to their protein class using PANTHER (v14.1). Subclassification of nucleic acid binding category is also shown. Number indicates the absolute number of genes in the category. ASE analyzed are all the significant RBPMS knockdown regulated events with $|\Delta PSI| > 10\%$.

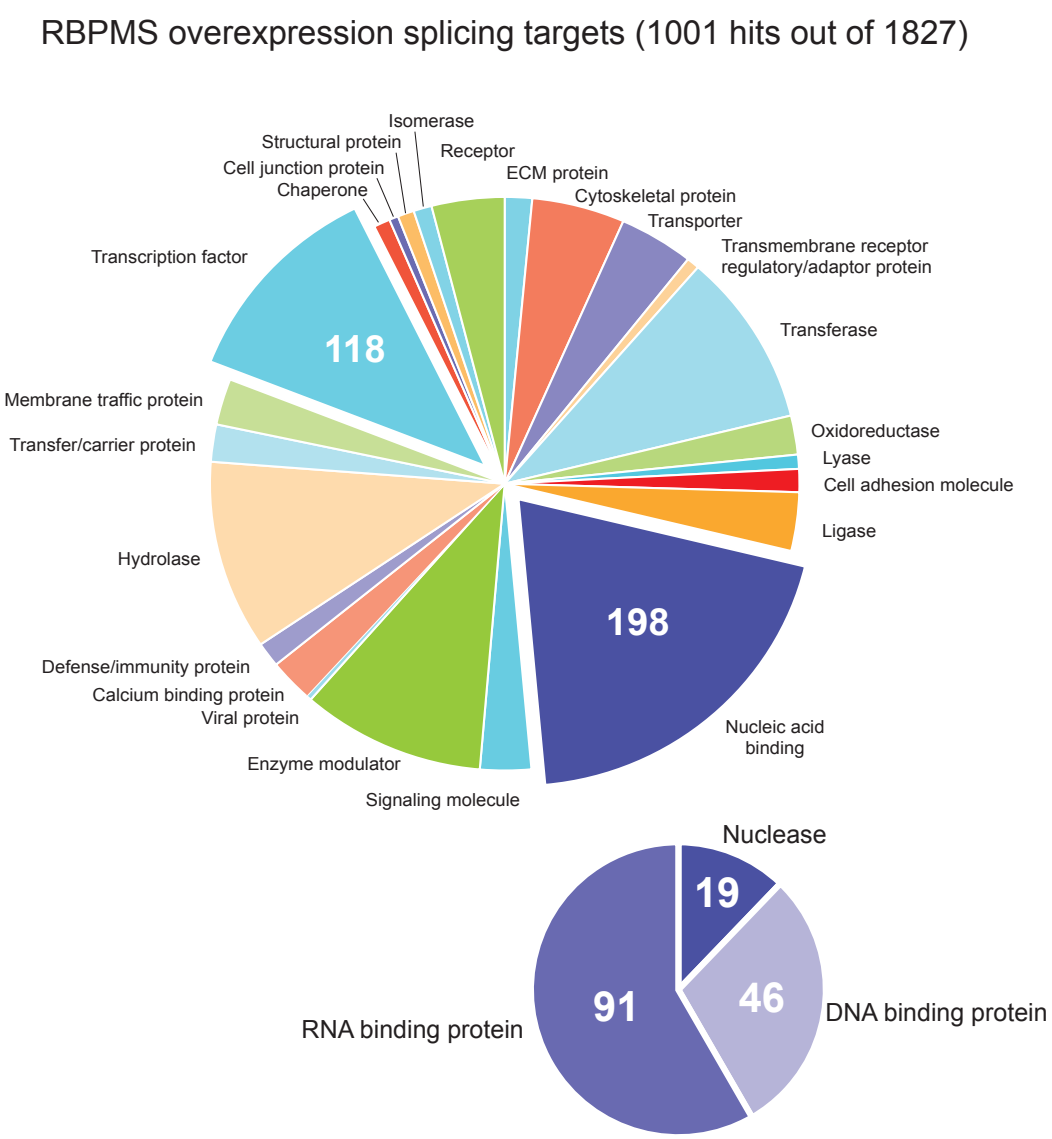


Fig. 6.5 RBPMS overexpression affects splicing of genes coding for a range of protein classes. Functional classification of RBPMS overexpression splicing targets according to their protein class using PANTHER (v14.1). Subclassification of nucleic acid binding category is also shown. Number indicates the absolute number of genes in the category. ASE analyzed are all the significant RBPMS overexpression regulated events with $|\Delta PSI| > 10\%$.

To get a more reliable and specific result, a second approach to investigate whether RBPMS controlled other regulators was applied using two curated lists of transcription factors and RBPs (Gerstberger et al., 2014; Lambert et al., 2018). These functional subset of genes found differentially spliced upon RBPMS knockdown and overexpression are shown in Table 6.1 and Table 6.2. In fact, in these analyses the lists of ASEs regulated by RBPMS used were the functional one comprising cassette exons commonly regulated between RBPMS knockdown and PAC1 dedifferentiation or RBPMS overexpression and aorta tissue dedifferentiation (see Material and Methods Chapter). Five post-transcriptional regulators were found regulated by RBPMS knockdown, with four of which were also regulated by RBPMS-A overexpression (Table 6.1). On the other hand, RBPMS-A overexpression affected a more extended list of regulators, including transcription factors (Table 6.2). In total it affected 20 genes with some of those containing multiple exons targeted by RBPMS, resulting in a list of 31 cassette exons differentially spliced (Table 6.2). Four genes were regulated in both experiments: *Lrrfip1*, *Mbnl1*, *Pstk* and *Rbm26*. However, for *Rbm26* the ASEs regulated in RBPMS knockdown and overexpression were different. Because *Mbnl1* was the most well studied RBP among the identified regulators, its splicing regulation by RBPMS was first chosen to be further characterized.

Table 6.1 Transcriptional (TF) and post-transcriptional (PT) regulators differentially spliced by RBPMS knockdown

RBPMS knockdown						
Gene	Chr	Strand	Exon Start	Exon End	ΔPSI	Factor type
Eif5a	10	-	56530568	56530662	17.0	PT
Lrrfip1	9	+	98243215	98243287	-22.6	PT
Mbnl1	2	+	150864085	150864121	37.4	PT
Pstk	1	+	201952551	201952846	-17.6	PT
Rbm26	15	-	89296794	89296866	-14.5	PT

Therefore, this paved the way for potential secondary targets of RBPMS by regulation of other post-transcriptional regulators. Consistent with the hypothesis of this study, regulation of regulators is a suggested feature of master splicing regulators (Jangi and Sharp, 2014).

Table 6.2 Transcriptional (TF) and post-transcriptional (PT) regulators differentially spliced by RBPMS overexpression

RBPMS overexpression						
Gene	Chr	Strand	Exon Start	Exon End	ΔPSI	Factor type
Akap17a	12	-	18530563	18530726	-44.3	PT
Akap17a	12	-	18529849	18530726	-34.9	PT
Ddit3	7	+	70582844	70582892	-33.2	TF
Dzip1	15	-	104170509	104170555	-21.4	TF
E2f5	2	-	88354476	88354612	-20.3	TF
Lrrfip1	9	+	98233128	98233206	23.0	PT
Lrrfip1	9	+	98243215	98243287	10.6	PT
Lsm14b	3	+	175415535	175415613	-72.6	PT
Lsm14b	3	+	175412593	175412710	-30.4	PT
Mbnl1	2	+	150864085	150864121	-43.5	PT
Mbnl1	2	+	150864801	150864896	-42.0	PT
Mbnl2	15	+	105772308	105772403	-40.7	TF/PT
Mrps18b	20	+	3348548	3348615	-42.6	PT
Mrps18b	20	+	3346556	3346625	-35.4	PT
Mrps18b	20	+	3346164	3346258	-19.2	PT
Mrrf	3	+	15475320	15475476	-13.9	PT
Mterf1	4	+	27366889	27367017	-23.6	TF
Mterf1	4	+	27367248	27367294	28.4	TF
Mybl1	5	+	9310670	9310850	28.4	TF
Myef2	3	-	117363827	117363878	-25.5	PT
Myef2	3	-	117362732	117362804	-20.0	PT
Papd7	1	+	36428872	36429031	-10.7	PT
Pprc1	1	+	265822338	265823056	-10.6	PT
Pprc1	1	+	265815951	265816044	-15.4	PT
Pprc1	1	+	265821381	265821511	-11.8	PT
Pstk	1	+	201952551	201952846	29.7	PT
Rbm26	15	-	89295925	89295997	-59.1	PT
Sp140	9	+	92657393	92657432	-19.4	TF
Stat1	9	-	54327359	54327491	-29.9	TF
Stat1	9	-	54327424	54327491	-34.6	TF
Tdrd7	5	+	61591965	61592183	-21.7	PT

6.2.2 RBPMS regulates splicing of *Mbnl1*, another splicing regulator

Muscleblind proteins have been implicated in regulation of AS of pre-mRNAs in health and disease (reviewed in Fernandez-Costa et al. (2011)), also being involved in other post-transcriptional processes such as mRNA localization (Wang et al., 2012), stability (Masuda et al., 2012) and alternative polyadenylation (Batra et al., 2014). Moreover, MBNL1 isoforms and paralogs have been shown to differ in their splicing activity and localization (Kino et al., 2015; Sznajder et al., 2016; Tabaglio et al., 2018; Tran et al., 2011). Therefore, regulation of MBNL1 isoforms by RBPMS could lead to broad downstream effects, including further secondary ASE in the SMCs.

Mbnl1 AS exons regulated by RBPMS were visualized by sashimi plots from RNAseq (Figure 6.6). Both RBPMS knockdown and overexpression in PAC1 cells indicated RBPMS as a repressor of the *Mbnl1* alternative exons referred to here as exons 7 and 8 and also specified by their sizes 36 and 95 nt respectively (Figure 6.6). Consistent with the RNAseq, the switch in RNA isoform was also significant in the RT-PCR validation (Figure 6.6). RBPMS knockdown promoted inclusion of 52.3% of exon 7 and 65.2% of exon 8, with an inclusion level difference of 33.6 and 16.6% between RBPMS knockdown and control (Figure 6.6). RBPMS-A overexpression also resulted in a clear switch as evident in the RNA-Seq data (Figure 6.6), and subsequently validated by C. Gooding (Appendix - Figure A.4). Furthermore, the change in the *Mbnl1* RNA isoforms was sufficient to alter the protein levels as seen in the western blot (Figure 6.6). However, not all the isoforms were identified at the protein level. Two main bands could be identified regulated by RBPMS in the western blots ∼35 and ∼43 kDa. Based on the expected size of the peptides encoded by exons 7 and 8, inclusion of each of these exons should result in a 1.3 or a 3.4 kDa shift respectively. Additionally, inclusion of both exons produces a larger MBNL1 isoform, ∼4.3 kDa protein shift. Thus, the two bands are likely to represent exon 8 only included isoform ($MBNL_8$) and both exons excluded isoform ($MBNL_{\Delta 7\text{-}8}$). The exact identity of the bands is hindered by additional splicing events in *Mbnl1* transcripts. Interestingly, RBPMS-A overexpression in PAC1 cells also regulated a *Mbnl1* paralog, *Mbnl2*, in which an alternative exon equivalent to exon 8 (95 nt) was found to be repressed upon RBPMS-A overexpression (Figure 6.7 and Appendix - Figure A.4). On the other hand, *Mbnl2* exon 7 was completely skipped in all the conditions (Appendix - Figure A.4). Thus, activity of both MBNL1 and 2 is likely to be decreased coordinately, a feature also observed in SM tissues in which alternative exons are more skipped.

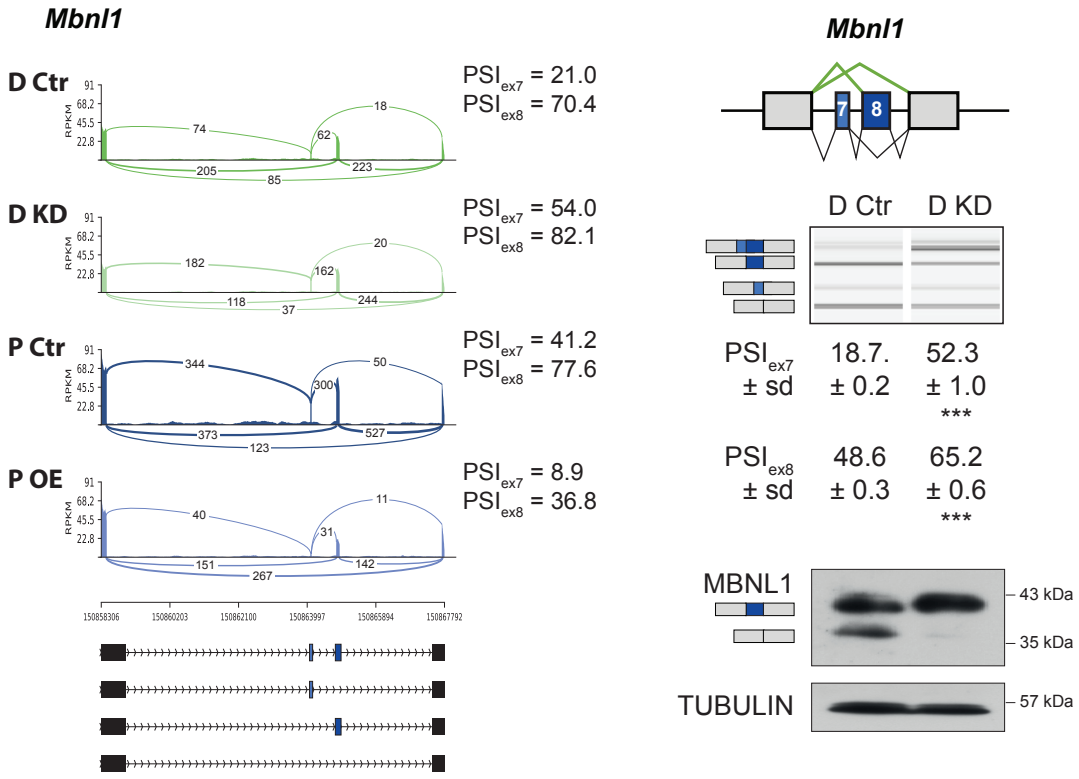


Fig. 6.6 RBPMS regulates *Mbnl1* alternative splicing. Left, sashimi plot of *Mbnl1* exons differentially spliced by RBPMS. The numbers on the top of the arches represent the number of exon-exon junction reads. PSI_{ex7} and PSI_{ex8} indicate the mean of the inclusion of exon 6 (36 bp) and 7 (95 bp) as calculated by rMATS. Schematic of the mRNA isoforms produced are found at the bottom as well as the chromosomal coordinates. **Right top**, RT-PCR validation of *Mbnl1* exon 7 and 8 by RBPMS knockdown in PAC1 cells. PSI values shown are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. **Right bottom**, western blot against MBNL1 with protein isoforms labeled on the left. TUBULIN was used as a loading control. Protein size markers are indicated on the right.

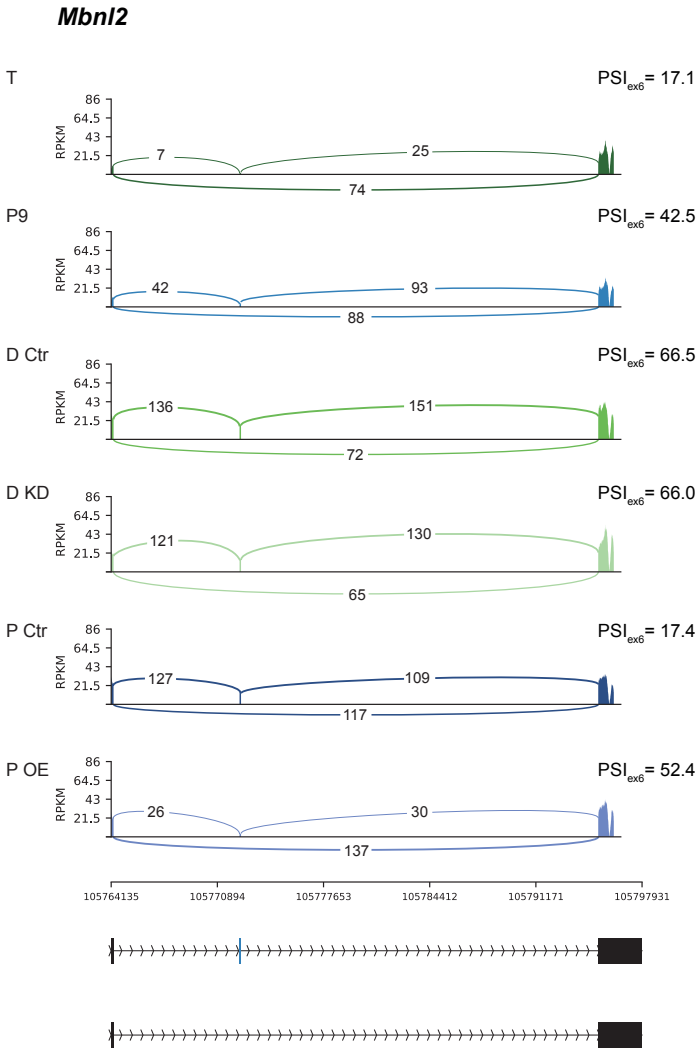


Fig. 6.7 RBPMS regulates *Mbnl2* alternative splicing. Sashimi plot of *Mbnl2* exons differentially spliced by RBPMS. The numbers on the top of the arches represent the number of exon-exon junction reads. PSI_{ex7} and PSI_{ex8} indicate the mean of the inclusion of exon 6 (36 bp) and 7 (95 bp) as calculated by rMATS. Schematic of the mRNA isoforms produced are found at the bottom as well as the chromosomal coordinates.

Exclusion of exon 7 and 8 of *Mbnl1* have previously been shown to affect MBNL1 localization and dimerisation, subsequently decreasing its splicing activity (Sznajder et al., 2016; Tabaglio et al., 2018). Thereby, it could be possible that some of RBPMS-regulated ASEs are in fact MBNL1 direct targets. To test this, C. Gooding performed MBNL1 and 2 knockdown while overexpressing RBPMS-A in proliferative PAC1 cells. In that manner, it was possible to distinguish ASE regulated by RBPMS, whose changes were also dependent on MBNL1. Indeed, *Ncor2* A5SS, previously identified by Sznajder et al. (2016) as an ASE that was less responsive to the shorter MBNL1 isoforms, was identified as an indirect target of RBPMS. *Ncor2* A5SS splicing was unresponsive to RBPMS overexpression when combined to MBNL1 and 2 knockdown (Appendix - Figure A.5). Nevertheless, *Ncor2* A5SS was detected to be significantly regulated by RBPMS-A overexpression by rMATS analysis, ΔPSI = 13%. In summary, RBPMS promotes shorter isoforms of MBNL1 and 2 that have previously been shown to be less active for some target ASEs (e.g. *Ncor2*).

6.2.3 RBPMS regulates splicing of *Lsm14b*, another post-transcriptional regulator

Next, we set to further characterize other potential post-transcriptional regulators affected by RBPMS. A promising candidate was *Lsm14b*, an RBP involved in the cytoplasmic control of mRNA stability and translation (Brandmann et al., 2018). *Lsm14b* alternative exon 6, which encodes for its only predicted NLS - RPPRRR, showed a very high inclusion level difference of 70% between the PAC1 cells control and RBPMS-A overexpression in the RNAseq analysis by rMATS (Table 6.2). Interestingly, *Lsm14b* is another gene whose splicing is regulated during aorta dedifferentiation but not in PAC1 cells as indicated in the Figure 6.8. In PAC1 cells, RBPMS-A overexpression recapitulated the tissue differentiated splicing pattern by decreasing the levels of inclusion of the *Lsm14b* exon to 11% (Figure 6.8). C. Gooding then confirmed the switch in isoforms at the RNA and protein levels by RT-PCR and western blot of the inducible RBPMS-A lentiviral populations (Appendix - Figure A.6). Despite not being significant in the RNAseq splicing analysis, RT-PCR analysis of *Lsm14b* AS upon RBPMS knockdown in PAC1 cells revealed a significant difference in the inclusion level of this cassette exon, 12% more included in knockdown samples (Figure 6.8). Furthermore, a more dramatic change was observed at the protein levels as indicated in the western blot (Figure 6.8). In fact, inclusion of the in-frame 29 amino acid peptide encoded by *Lsm14b* exon 6 is expected to produce a 3.1 kDa shift in

protein size (Figure 6.8). Finally, apart from its cytoplasmic roles, *Lsm14b* can also shuttle to the nucleus in which its function is still unclear (Brandmann et al., 2018). Thus, RBPMS mediated AS of *Lsm14b* could affect its localization and consequently also its mRNA turn over function.

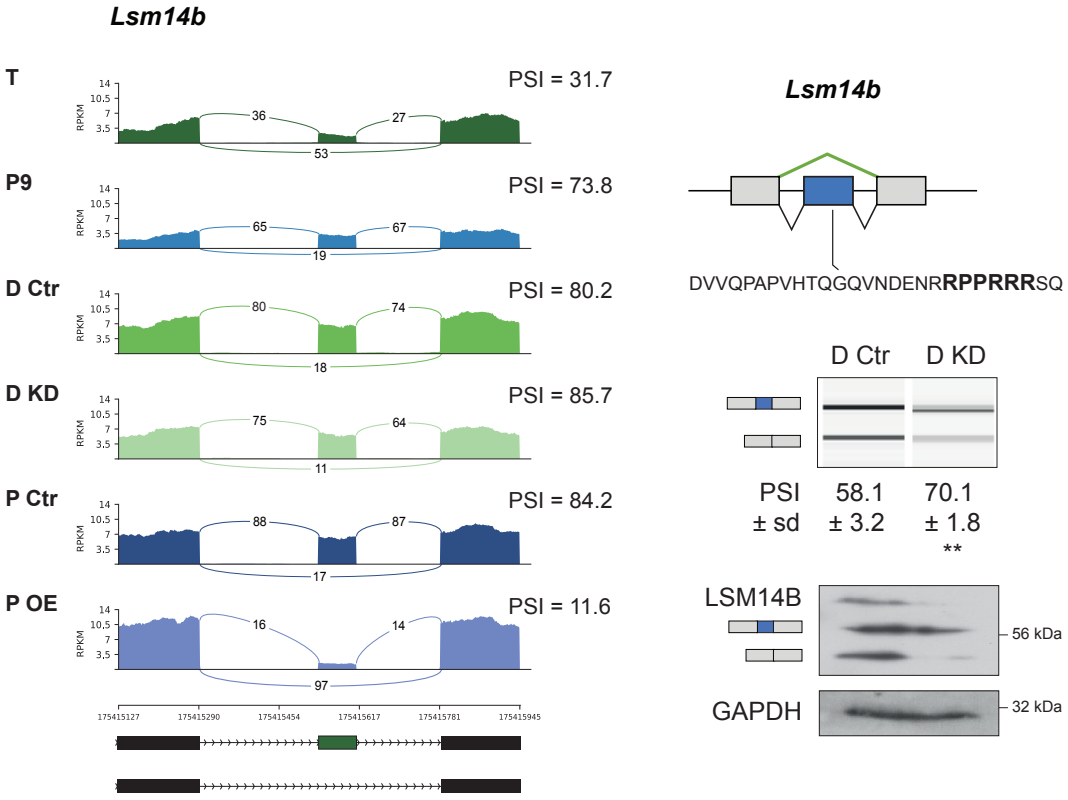


Fig. 6.8 *Lsm14b* a post-transcriptional regulator regulated by RBPMS. Left, sashimi plot of *Lsm14b* cassette exon differentially spliced upon RBPMS overexpression in PAC1 cells. Plots are shown for aorta dedifferentiation (tissue (T) and cultured primary SMC passage 9 (P9)), RBPMS knockdown and overexpression. The numbers on the top of the arches represent the number of exon-exon junction reads. PSI indicate the mean of the inclusion levels calculated by rMATS. Chromosomal coordinates and splicing isoforms are shown at the bottom. **Right**, schematic of *Lsm14b* splicing event. Amino acid encoded by the alternatively spliced exon is shown and the predicted NLS highlighted in bold font. RT-PCR validation in RBPMS knockdown. PCR products are indicated on the left. PSI values shown are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Western blot against LSM14B. GAPDH was used as a loading control. Protein size markers are shown on the right.

6.2.4 RBPMS regulates splicing of *Myocd*, a critical SMC transcriptional factor

Detailed inspection of transcription factors affected by RBPMS knockdown and overexpression in PAC1 cells only highlighted minor splicing isoforms in genes whose function in SMC biology are not clear (Table 6.2). Moreover, those changes were only observed in the overexpression dataset. Therefore, being aware of the SMCs and cardiac cells master transcriptional regulator, *Myocd*, we focused on any potential ASE in its gene that could be regulated by RBPMS. In fact, *Myocd* is found expressed as two major protein isoforms that arise from regulation of its alternative exon 2a (Creemers et al., 2006; van der Veer et al., 2013). In cardiac cells and proliferative SMCs, a full length MYOCD protein is produced by skipping of exon 2a (Creemers et al., 2006; van der Veer et al., 2013). On the other hand, in differentiated SMCs, exon 2a is mainly included, introducing an in-frame stop codon that then leads to the expression of an N-terminal truncated MYOCD isoform by alternative translation initiation from a downstream AUG codon in exon 4 (Creemers et al., 2006; van der Veer et al., 2013). Thereby, the SMC isoform lacks the N-terminal MEF2 interacting domain, enabling activation of SMC-specific promoters (Creemers et al., 2006; van der Veer et al., 2013) (Figure 6.9). Indeed, supporting a potential role of RBPMS in the regulation of this key SMC transcription factor, two conserved clusters of CACs were identified ∼200 nt downstream of *Myocd* exon 2a (Figure 6.10). The location of these binding sites suggested an activation role for RBPMS, in agreement with the promotion of the differentiated SMC specific isoform.

6.2.4.1 RBPMS promotes the inclusion of the *Myocd* SMC differentiated exon

Although rMATS analysis did not identify significant changes in the inclusion of *Myocd* exon 2a, manual inspection of RNAseq datasets, and subsequent RT-PCR analysis, revealed that *Myocd* exon 2a was more included in differentiated than proliferative PAC1 cells (Figure 6.11). Moreover, RBPMS knockdown decreased inclusion of exon 2a (Figure 6.11). Indeed, the inclusion levels calculated from the RNAseq manual inspection was in good agreement with the RT-PCR observed PSI (70.8% and 64.9% for differentiated PAC1 and 37.7% and 44.3% for RBPMS knockdown). In order to better understand exon 2a regulation, a minigene reporter containing *Myocd* exon 2a and its flanking intronic regions was created (Figure 6.11). PAC1 cells were then transfected with the minigene reporter, showing a basal inclusion level of ∼ 30%, in

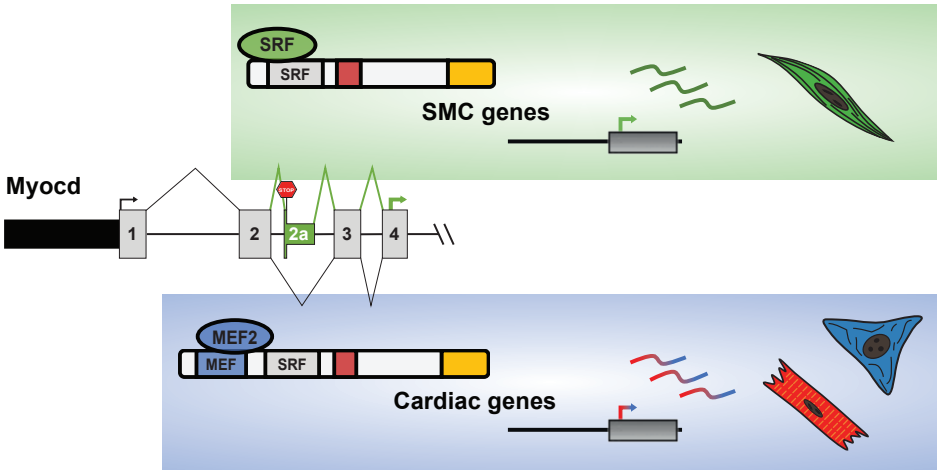


Fig. 6.9 *Myocd* exon 2a splicing in VSMC. Illustration of Myocd exon 2a ASE in SM and cardiac cells. MYOCD full length containing the MEF domain is produced by exclusion of exon 2a in cardiac and proliferative cells (blue). On the other hand, inclusion of exon 2a in SMCs, inserts a premature stop codon, leading to the translation of a downstream AUG codon, which codes for a shorter isoform that lacks the MEF domain (green). The absence of the MEF2 interacting domain subsequently affects the subset of genes activated by MYOCD.

the reporter only and Venus co-transfected controls (Figure 6.11). Expression of either RBPMS-A or B significantly promoted inclusion of *Myocd* exon 2a in the PAC1 cells to ∼70% (Figure 6.11). Thus, RBPMS knockdown and overexpression in PAC1 cells regulated inclusion of *Myocd* exon 2a.

6.2.4.2 RBPMS regulates *Myocd* exon 2a via binding to downstream CACs

Further investigation of the regulation of *Myocd* exon 2a by RBPMS was then achieved by transfecting HEK293 cells with CAC mutated minigenes reporters (Figure 6.12A-B). First, the wild-type minigene was tested in HEK293 and, as expected, exon 2a was not included in these non-smooth muscle cells (Figure 6.12C). However, transient overexpression of rat RBPMS-A and RBPMS-B promoted the inclusion of exon 2a, increasing the PSI to ∼75% and 37% respectively (Figure 6.12C). Secondly, to address the importance of the downstream CAC motifs, each cluster of CAC was mutated to CCC and tested in HEK293 cells with co-transfection of RBPMS-A expression vector (Figure 6.12C). Cluster 1 mutation (mCAC1) had minor effects on RBPMS-A response, 67% exon 2a inclusion, yet activation by RBPMS-B was more affected by

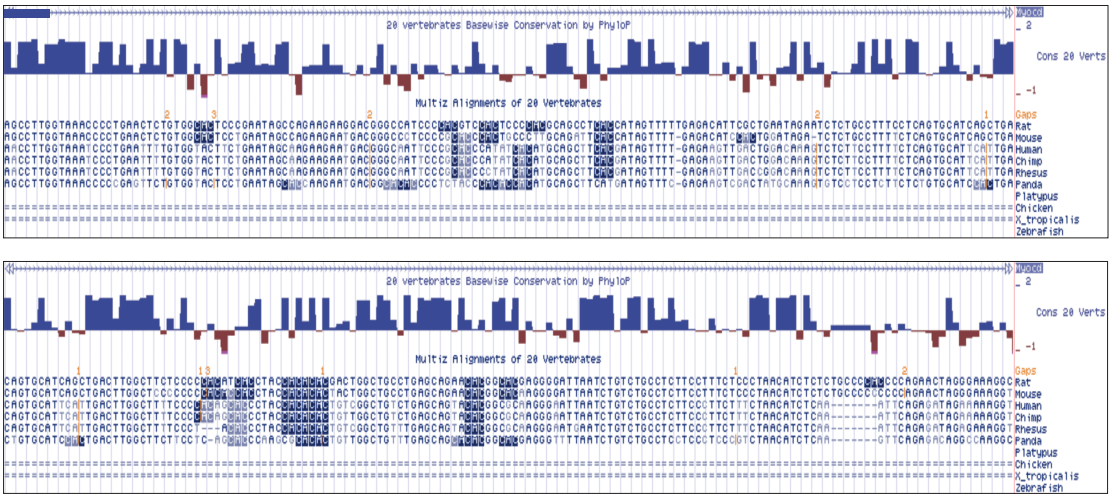


Fig. 6.10 Conserved CAC trinucleotides downstream of *Myocd* exon 2a. Conserved CAC motifs downstream of *Myocd* exon 2a are highlighted in blue. 3′ end of *Myocd* exon 2a is indicated by the purple box. Basewise conservation track and multiple alignment were generated by PhyloP and Multiz Alignments in the UCSC genome browser. Chr coordinates from *rn6*: top panel, chr10:51,733,829-51,733,977, and, lower panel, chr10:51,733,695-51,733,843

this cluster, reducing the inclusion level of exon 2a to 8.6%. Mutation of the second cluster (mCAC2) affected regulation by both isoforms, decreasing the inclusion by ∼2 fold. But, it was the combined mutant (mCAC) that had the strongest effect in the regulation by RBPMS, making the *Myocd* exon 2a reporter completely unresponsive to RBPMS expression (Figure 6.12C).

RBPMS binding to *Myocd* RNA was investigated by *in vitro* binding assays using both EMSA and UV-crosslinking. To confirm binding to the CAC clusters, assays were performed with wild-type (wt) and CAC mutants (mCAC) using *in vitro* transcribed RNAs (Figure 6.12D). In EMSA, both RBPMS isoforms bound to the wild-type RNA at similar apparent affinities ($0.125\mu M < K_d < 0.5\mu M$) (Figure 6.12D). *Myocd* mCAC1 and mCAC2 were associated with more modest ($K_d = \sim 0.5\mu M$) and drastic effects ($K_d = \sim 2.0\mu M$) respectively. Additionally, mutation of both CAC clusters (mCAC reporter) completely impaired binding by RBPMS. UV-crosslinking of *Myocd* RNA substrates and recombinant RBPMS was found to be very inefficient, with very faint signal observed only for the wild-type and to a less degree to the mCAC1 (Figure 6.12D lower panel). Therefore, consistent with the splicing experiments, the *in vitro* binding assays indicate that RBPMS binds to the CAC motifs downstream of *Myocd* exon 2a, with these sites being critical for its splicing regulation. Moreover, individual

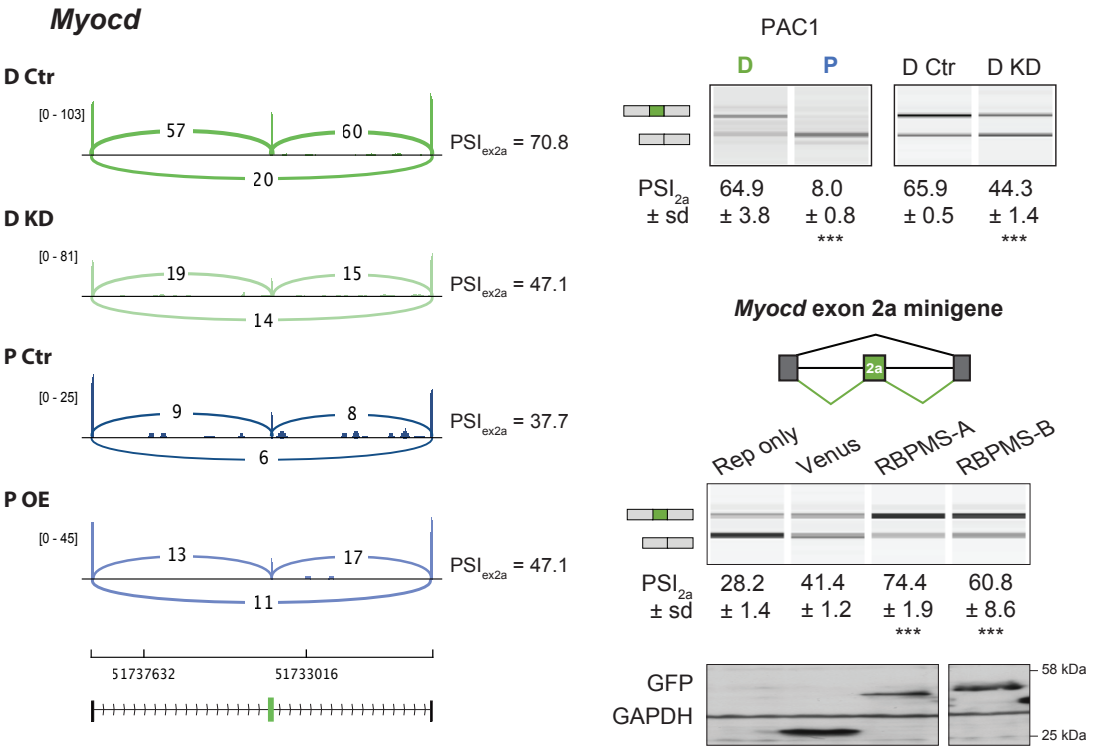


Fig. 6.11 RBPMS promotes the splicing of the differentiated *Myocd* isoform in PAC1 cells. Left, Sashimi plot of *Myocd* cassette exon differentially spliced upon RBPMS overexpression in PAC1 cells. Numbers on top of the arches represent the number of exon-exon junction reads. PSI indicate the mean of the inclusion levels calculated by rMATS. Chromosomal coordinates and splicing isoforms are shown at the bottom. **Right top**, RT-PCR validation of the SMC pattern of *Myocd* exon 2a in the PAC1 dedifferentiation and RBPMS knockdown in PAC1 cells. **Right bottom**, schematic of the rat *Myocd* exon 2a minigene. Splicing of the *Myocd* exon 2a minigene reporter in PAC1 cells upon transient transfection of RBPMS-A and RBPMS-B. Reporter only and Venus controls were tested alongside. Western blot against GFP and, GAPDH was used as a loading control. Protein size markers are shown on the right. PCR products are indicated on the left. PSI values shown are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

mutation of each CAC cluster, mCAC1 and mCAC2, despite reducing the intensity of the binding (Figure 6.12), did not provide evidence of an intermediate size complex. Given that each cluster has sufficient CAC motifs to bind at least one dimer this suggests that RBPMS might be binding to RNA as a higher order oligomer than just a dimer.

6.2.4.3 Insertion of previously defined RBPMS site rescues RBPMS regulation of mCAC *Myocd*

In an attempt to rescue RBPMS activation of *Myocd* exon 2a, a sequence from the *UBE2V1* UTR, previously defined by PAR-CLIP to bind to RBPMS-A, was introduced into the *Myocd* mCAC reporter (Farazi et al., 2014). The sequence containing three CAC motifs separated by 1 and 3 nt was inserted into both locations of the mutated *Myocd* CAC clusters (Figure 6.13). Regulation of exon 2a was then tested in HEK293 cells. In agreement with the last experiment, *Myocd* double mutant was not able to promote inclusion of exon 2a in response to RBPMS-A or B. Nevertheless, insertion of RBPMS sites from *Ube2v1* significantly restored activation of exon 2a by RBPMS-A (PSI = 31.7%), with minor effects upon RBPMS-B overexpression (PSI= 6.8%), despite similar levels of protein expression of RBPMS isoforms shown by the western blots (Figure 6.13). Moreover, the RBPMS sites from *Ube2v1* did not affect the basal level of inclusion of exon 2a in HEK293 cells as indicated by the reporter only and Venus controls (Figure 6.13). This is consistent with the expected *Myocd* splicing pattern in a non-smooth muscle cell. Thus, exchanging the CAC motifs with a predefined RBPMS site in the *Myocd* minigene reporter reproduces the regulation by RBPMS-A, yet the reason behind the isoform specific rescue remains unclear.

6.2.5 RBPMS and QKI play antagonistic roles in VSMC

QKI is another RBP that has been established as important in the SMC dedifferentiation process, regulating the proliferative phenotype, the state in which it is found highly expressed (Figure 3.5) (Llorian et al., 2016; van der Veer et al., 2013). Moreover, QKI was also reported to be able to control the SMC dedifferentiated state by controlling splicing of *Myocd* exon 2a (van der Veer et al., 2013). QKI binds to the 5′ end of the exon 2a repressing its inclusion in proliferative SMCs (van der Veer et al., 2013). Therefore, RBPMS and QKI could potentially play antagonistic roles in the regulation of the SMC AS during the phenotypic switch of these cells. To test the functional antagonism in the regulation of exon 2a, HEK293 cells were co-transfected with *Myocd*

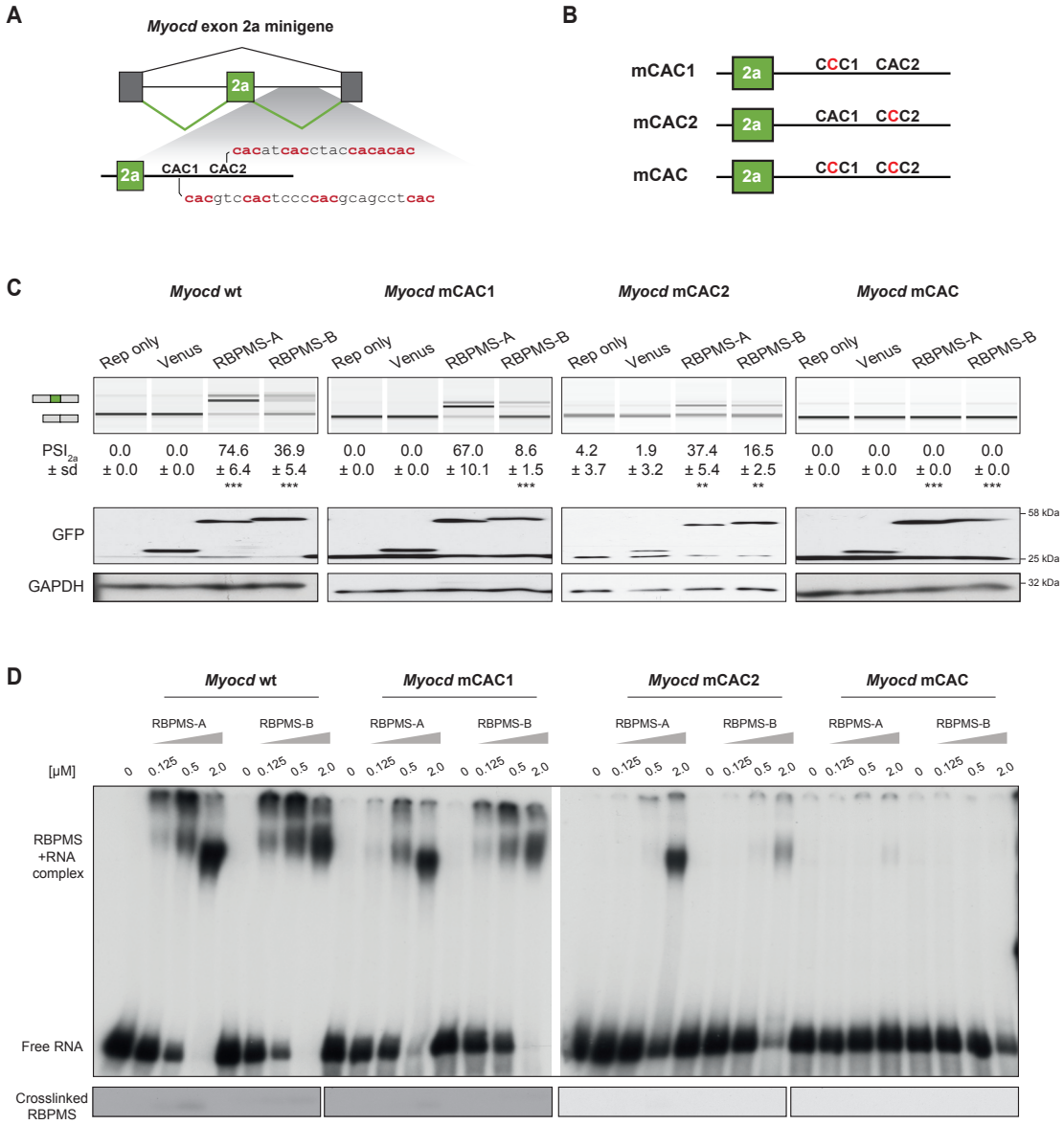


Fig. 6.12 RBPMS regulates splicing of *Myocd* exon 2a via downstream CACs. **A** Schematic of the *Myocd* exon 2a minigene with downstream clusters of CACs highlighted. **B** Schematic of the *Myocd* minigene mutant reporters. **C** RT-PCRs to verify the splicing patterns of *Myocd* exon 2a reporters (wild-type and mutants) upon RBPMS-A and RBPMS-B overexpression in HEK293 cells. Reporter only and Venus controls were tested in parallel. PCR products are indicated on the left. PSI values shown are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Protein expression levels are shown in the western blot probing for GFP and GAPDH as a loading control. Protein size markers are shown on the right. **D** *in vitro* binding of RBPMS-A and RBPMS-B to *Myocd* exon 2a CAC clusters was tested by EMSA, top, and UV-crosslinking, bottom. *in vitro* transcribed wild-type and CAC mutant radiolabelled RNAs were used. Binding by recombinant RBPMS was tested in a serial dilution (1:4) in a range of 0.125 to 2 μM.

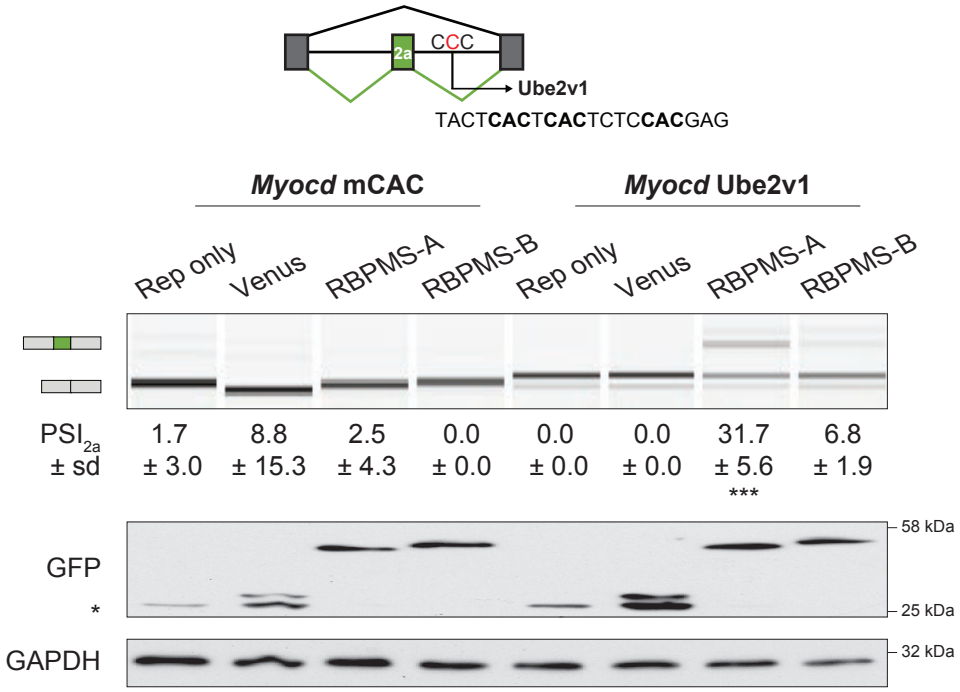


Fig. 6.13 Inclusion of exon 2a in mCAC *Myocd* reporter can be rescued by insertion of a defined RBPMS site. Top, a previously defined RBPMS binding sequence from *UBE2V1* (Farazi et al., 2014) was inserted into the mCAC *Myocd* minigene and its splicing tested upon RBPMS overexpression in HEK293 cells. UBE2V1 sequence is shown in the *Myocd* minigene schematic. **Middle**, RT-PCR were performed including reporter only and Venus controls. PCR products are indicated on the left. PSI values shown are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. **Bottom**, western blot probing for GFP and GAPDH as a loading control. Asterisk on the side of the gel indicates the GFP by-product from the minigene splicing.

reporter and both RBPs, RBPMS and QKI (Figure6.14). RBPMS-A and RBPMS-B promoted inclusion of *Myocd* exon 2a (88.4% and 67.7%), whereas expression of QKI repressed its inclusion (PSI= 0.0%). When co-expressing the two RBPs, QKI even at low concentrations strongly antagonized activation by RBPMS-A and RBPMS-B (PSI< 10%), repressing exon 2a below the basal level observed in HEK293 cells as indicated by the Venus control (PSI= 17.5%). Thus, these data show that *Myocd* exon 2a is under the control of RBPMS in the differentiated SMC and of QKI in the proliferative SMCs, with QKI being the dominant regulator in this antagonistic regulation.

Interestingly, *Flnb* exon H1, which I found to be activated by RBPMS (Chapters 3-5), was recently shown to also be regulated by QKI during the epithelial mesenchymal transition (Li et al., 2018). Similar to regulation of *Myocd*, QKI repressed inclusion of exon H1, but in this case, as revealed by eCLIP (enhanced CLIP), repression was mediated by unusual binding of QKI downstream of the regulated exon, which is usually associated with activation (Li et al., 2018). To test for functional antagonism in *Flnb* as well, endogenous splicing of exon H1 was verified in the previous co-transfection with RBPMS and QKI and *Myocd* reporter (Figure 6.14). Consistent with the *Myocd* results, QKI expression decreased inclusion of exon H1 by RBPMS isoforms from 50.6% to 22.8% for RBPMS-A and from 56.5% to 11.6% for RBPMS-B. Nevertheless, the reduction was not as drastic as the one observed in the regulation of *Myocd* exon 2a. Thus, splicing of *Flnb* exon H1 is also under control of RBPMS and QKI with no clear dominant regulator in this splicing event. In addition to that, QKI motifs were found enriched in RBPMS splicing targets as shown in Figure6.15. QKI binding sites (ACTAACAA), were mainly distributed around exons more included by RBPMS. Together, these data suggest a potential co-regulation of the SMC splicing by these two RBPs.

6.3 Discussion

6.3.1 RBPMS regulates activity of other regulators via splicing

In addition to regulation of various mRNAs involved in cytoskeleton and cell adhesion, RBPMS also mediated alternative splicing changes in transcriptional and post-transcriptional regulators. By regulating MBNL, LSM14B and MYOCD, RBPMS has the potential for more widespread actions (Figures 6.6, 6.8 and 6.12).

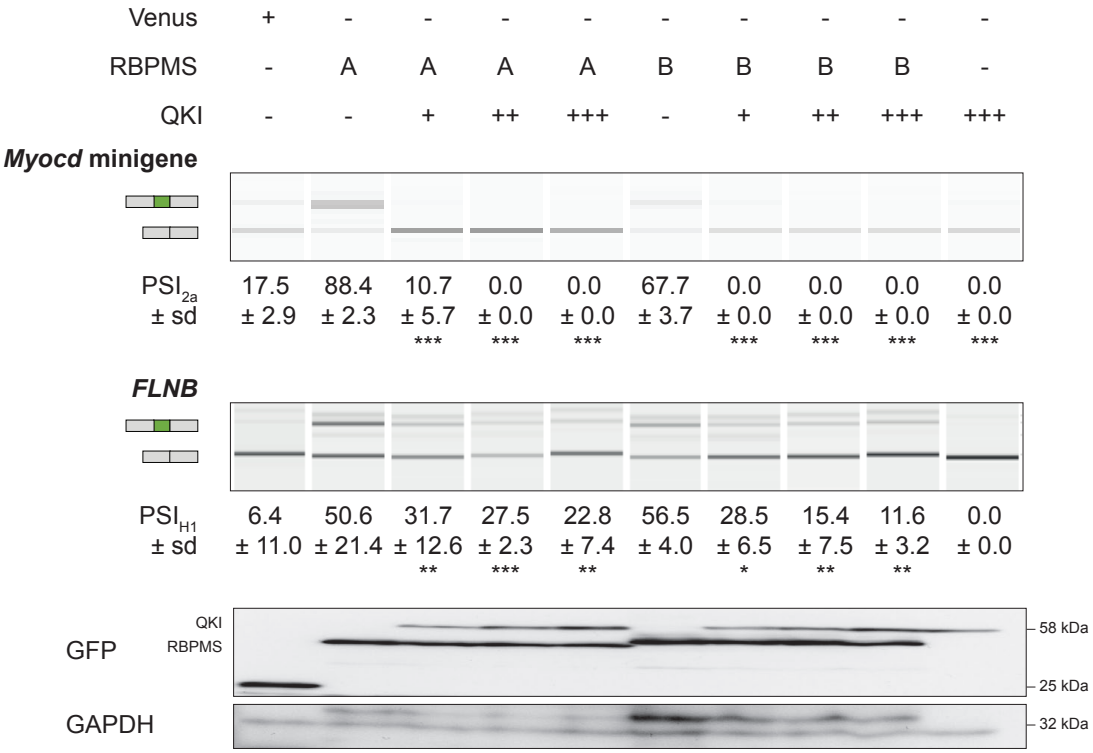


Fig. 6.14 RBPMS and QKI regulate similar targets in SMC. Splicing patterns of rat *Myocd* exon 2a minigene reporter and *FLNB* exon H1 upon co-expression of RBPMS and QKI in HEK293 cells. PCR products are identified by schematics on the left. PSI values shown are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Western blot against GFP to verify overexpression of Venus tagged RBPMS and QKI, respectively labeled on the left. GAPDH was used as a loading control. Protein size markers are indicated on the right. +, ++ and +++ indicate QKI increasing concentration in the co-expression.

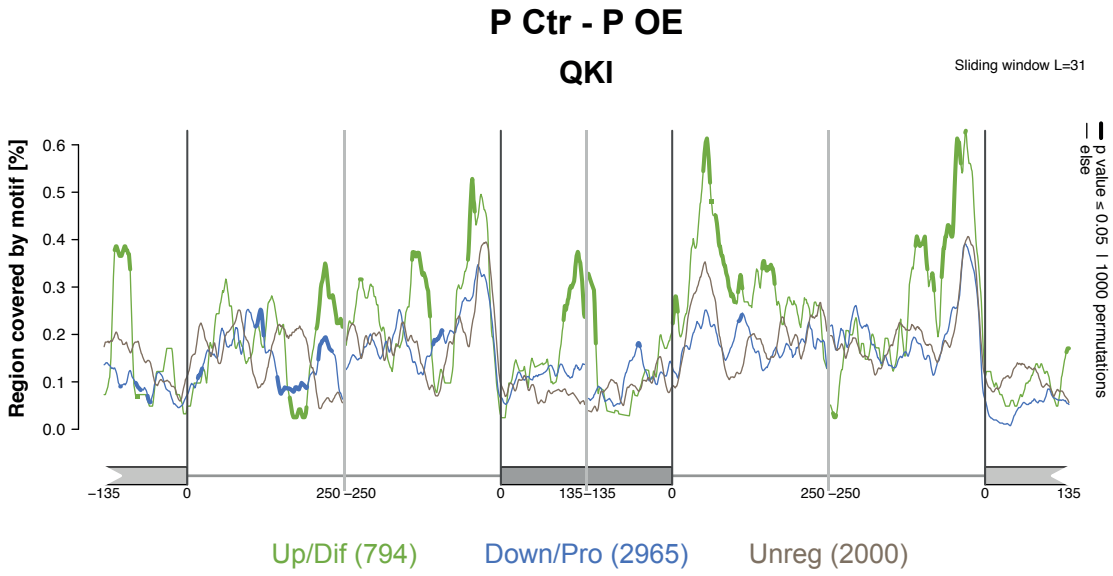


Fig. 6.15 RBPMS splicing targets are enriched for QKI binding sites (ACTAACAA). QKI motif maps in regulated cassette exons from RBPMS overexpression (P OE - P Ctr) in PAC1 cells. Motif enrichment as well as depletion is shown for upregulated or differentiated exons (green), downregulated or proliferative exons (blue) and unregulated exons (gray). Motif maps were generated using the Matt toolkit. Statistical significance calculated by the Matt toolkit is shown by the line width.

First, RBPMS affected ASEs in MBNL proteins that have direct effects on their splicing regulatory activities (Sznajder et al., 2016; Tabaglio et al., 2018; Tran et al., 2011). Despite the limitation of the short-term overexpression experiment for the detection of secondary AS changes, a modest change in *Ncor2* A5SS seems to reflect an RBPMS-induced switch to less active MBNL isoforms (Appendix - Figure A.5, validated by Clare Gooding in the laboratory). Implications of RBPMS-mediated splicing of *Mbnl1* and *Mbnl2* are summarized in Figure 6.16.

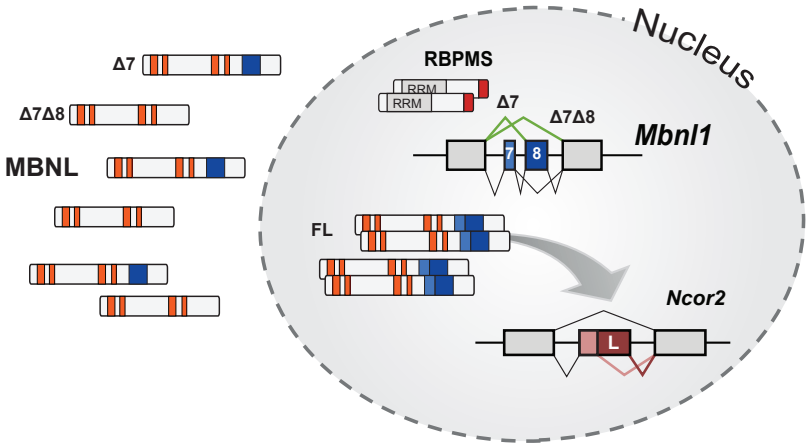


Fig. 6.16 RBPMS promotes the production of less active *Mbnl1* isoforms. Schematic of RBPMS indirect regulation of *Ncor2* splicing via control of MBNL1 isoforms.

RBPMS also affected another RBP involved in non-splicing functions, *Lsm14b*. The regulated *Lsm14b* event could potentially modulate mRNA stability, or an as yet uncharacterized LSM14B role in the nucleus (Kırlı et al., 2015). Potential implications of RBPMS-mediated splicing of *Lsm14b* is summarized in Figure 6.16

Lastly, RBPMS also directly targeted a critical SMC transcription factor, *Myocd*. Interestingly, a similar regulation has been attributed to the myogenic AS master regulator RBM24 (Jangi and Sharp, 2014), which is important for stabilization of the transcription factor *Myogenin* mRNAs by binding to its 3′UTR (Jin et al., 2010). RBPMS direct activation of *Myocd* exon 2a drives production of a more potent isoform of the VSMC transcription factor which is associated with more differentiated SMCs (van der Veer et al., 2013).

Additional effects upon transcription could also be conferred by RBPMS activation of the *Flnb* H1 exon. Inclusion of the exon H1 produces a more nuclear FLNB isoform which antagonizes the transcription factor FOXC1 in epithelial cells (Li et al., 2018).

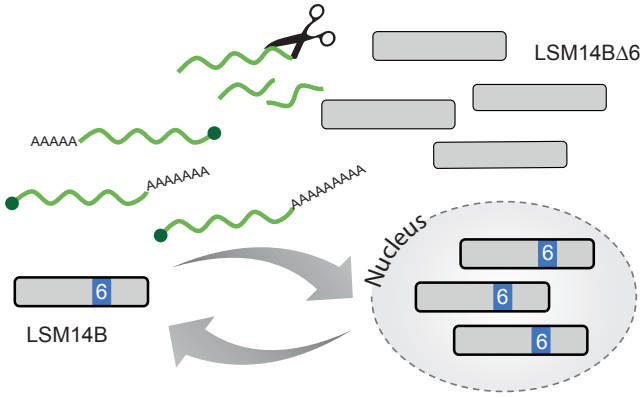


Fig. 6.17 Potential consequence of *Lsm14b* alternative splicing by RBPMS. Schematic of the impairment of LSM14B nuclear shuttling and function in mRNA turnover by RBPMS regulation of *Lsm14b* exon 6 which contains the only predicted nuclear localization signal (shown in blue).

Indeed, FOXC1 and FOXC2 are expressed at higher levels in adult arteries than any other human tissue (Lonsdale et al., 2013). Thus, RBPMS activation of this ASE primarily affect SMCs by altering the cytoskeleton architecture but it could also lead to additional effects upon transcription via regulation of FOXC1. In agreement with a role in SMC biology, FOXC1 and FOXC2 are critical for the early remodeling of blood vessels (Kume et al., 2001).

In Jangi and Sharp (2014), it was suggested that master regulators could regulate other splicing factors via AS-NMD. However, instead of regulating splicing events controlling the abundance of other regulators, RBPMS regulated post-transcription and transcription factors by production of isoforms with differential activity, similarly to the dominant negative isoform generated by RBFOX proteins (Damianov and Black, 2010). Additionally, RBPMS-mediated splicing of *Lsm14b* outlines another manner by which RBP functions can be controlled, which is by altering their subcellular localization. A well characterized example of that is the one represented by the QKI protein isoforms (Fagg et al., 2017). All of these processes comprise different mechanisms beyond AS-NMD that assist in the establishment of robust and cell-specific transcriptomes.

Another interesting possibility is that RBPMS might affect other RNA processes such as miRNA biogenesis. Recent studies have already uncovered the role of a number of splicing factors in the miRNA processing (Michlewski and Cáceres, 2019). In fact, RBPMS could potentially regulate the SMC miRNAs, miR143-145 (Boettger et al., 2009; Cordes et al., 2009), as suggested by the presence of CAC motifs in their precursor

miRNA. Thus, control of miRNAs could consist of an extra layer of regulation of the SMC transcriptome and biology by RBPMS.

6.3.2 RBPMS oligomeric state

Band shift assays using recombinant RBPMS-A and B proteins and *in vitro* transcribed *Myocd* RNA substrates with individual CAC clusters mutated (mCAC1 and mCAC2), in addition to the reduction of binding, also revealed an interesting feature regarding the size of the complex RNA-RBPMS formed (Figure 6.12). Each cluster contains enough motifs to accommodate at least one dimer (potential four binding sites). Therefore, smaller intermediate complexes were expected to be formed upon mutation of each CAC cluster *Myocd*. Nevertheless, RBPMS-RNA complexes of the same size of the wt RNA were observed for both isoforms. Furthermore, previous EMSA assays with *Flnb* and *Tpm1* also formed large complexes that required optimization of gel conditions to allow complexes to enter the gel. Thus, it is likely that RBPMS binds to RNA in higher order assembly, a property that might be common to both isoforms as well as to repressive and activated RNAs. Consistent with our data, Farazi et al. (2014) reported a molecular weight higher than a monomer for recombinant RBPMS full-length when investigating the oligomeric state of RBPMS in the absence of RNA by gel filtration. In fact, RBPMS oligomeric state could not be resolved due to the fact that the full-length FLAG-HA-tagged protein elution volume exceeded the resolution limit of the gel-filtration column (Farazi et al., 2014). However, N-terminal and C-terminal truncated RBPMS, as well as the combined deletion mutant, resolved to a size corresponding to a tetramer. Thereby, this assembly of larger assemblies is probably driven by additional sequences to the RRM domain of RBPMS and may be affected by the different isoforms. RBPMS structural and functional studies are presented in Chapter 7.

6.3.3 RBPMS and QKI in the establishment of differentiated and proliferative SMC programs

Cell-specific splicing programs usually result from the combination of multiple regulatory RBPs. QKI, PTBP1 and SRSF1 are all RBPs identified to promote the proliferative SMC phenotype (Llorian et al., 2016; van der Veer et al., 2013; Xie et al., 2017). In here, RBPMS and QKI antagonistically regulated at least two targets (*Myocd* and *Flnb*) (Figure 6.14). Another potential ASE co-regulated by these two RBPs is the penultimate exon of *Smtn*, which is repressed by RBPMS but activated by QKI (Llorian et al., 2016). In agreement with that, QKI motif enrichment analysis pointed

out significant association of QKI binding sites in RBPMS regulated exons (Figure 6.15). Thus, it is likely that RBPMS and QKI target a common set of ASEs but act with opposing roles, regulating the differentiated and proliferative SMC splicing programs. Indeed, the expression of the two RBPs are reciprocally regulated during SMC dedifferentiation, indicating that the switch-like changes observed in many ASE could be a consequence of their antagonistic activity.

Therefore, consistent with its designation as a master splicing regulator, RBPMS regulated other post-transcriptional regulators and directly promoted VSMC-specific splicing of the transcription factor Myocd by activating exon 2a inclusion, which differentiates SMC and cardiac isoforms and promotes the contractile VSMC phenotype. Moreover, identification of QKI co-regulated ASEs provided insights into potential interplay between the two RBPs in the establishment of contractile or synthetic SMC splicing programs.

6.4 Final conclusions

In conclusion, this chapter aimed to determine whether among the direct targets of RBPMS there were events controlling other RBPs leading to secondary splicing changes as well as other post-transcriptional processes. In addition, splicing regulation of transcriptional factors was also addressed. Inspection of RBPMS regulated events led to the identification of *Mbnl1* and *2*, *Lsm14b* and *Myocd* as targets that could drive to secondary effects in mRNA splicing, stability and gene expression. Therefore, compatible with its role as a master splicing regulator, RBPMS potentially increases the robustness of its splicing network in SMC by integrating other RBPs and transcription factors. Furthermore, QKI was shown to strongly antagonize the splicing of at least two RBPMS targets, *Myocd* and *Flnb*, suggesting an interplay between these two RBPs in SMC dedifferentiation.

In summary, the main findings of this chapter are:

1. RBPMS regulates splicing of *Mbnl1* and *2* generating less active isoforms that lead to secondary splicing changes.

2. RBPMS affected *Lsm14b* by repressing its exon 6 which encodes a predicted NLS.

3. RBPMS directly controlled the critical SMC transcription factor *Myocd* via binding to downstream CAC motifs.

4. Insertion of an RBPMS-defined site rescued inclusion of *Myocd* exon 2a by RBPMS-A.

5. RBPMS and QKI proteins antagonistically regulated *Myocd* exon 2a and *Flnb* exon H1.

In addition to that, this chapter also led to other questions listed below:

- Which targets are directly regulated by RBPMS?
- Are there other indirect targets among RBPMS regulated splicing events?
- Does *Mbnl* isoform switch control other ASEs?
- Does *Lsm14b* exon 6 affect its subcellular localization? If it does, what is the role of LSM14B in the nucleus as well as which transcript it targets?
- Does regulation of *Myocd* exon 2a by RBPMS lead to phenotypic changes in the SMCs?
- Do further *cis*-element features determine differential binding by RBPMS isoforms?
- Do RBPMS and QKI form a general antagonistic mechanism for establishment of switch-like splicing?
- Do other RBPs compete or cooperate in the regulation of RBPMS targets?

These questions are further discussed as potential future studies in the final chapter of this thesis.

Chapter 7

RBPMS: anatomy of a regulator

7.1 Introduction

7.1.1 RBPMS domains and structure

The RBPMS protein family comprises RBPs that contain a single RRM (Farazi et al., 2014). Conversely to RBPMS, most other RRM-containing proteins usually show this motif repeated in tandem or together with other RNA-binding domains (Daubner et al., 2013; Lunde et al., 2007). Nevertheless, Rbfox proteins (RBFOX1, RBFOX2, RBFOX3) also contain a single RRM which binds RNA and do not dimerize (Auweter et al., 2006), yet QKI proteins, also single RNA-binding domain, KH, containing RBPs, are found forming homodimers (Beuck et al., 2012). In fact, the structure of both human RBPMS and RBPMS2 RRMs have been solved revealing their RNA binding and dimerization interface (Sagnol et al., 2016; Teplova et al., 2016). The structures provided insights into RBPMS recognition of RNA at the molecular level as discussed in Chapter 5 (Sagnol et al., 2016; Teplova et al., 2016). Furthermore, RBPMS RRM has been shown to drive its symmetrical dimerization in the free state, with CAC motifs specifically bound by each monomer (Farazi et al., 2014) (Figure 7.1). RBPMS dimeric state is consistent with its PAR-CLIP identified RNA recognition element comprising tandem CAC motifs separated by variable spacer length (Farazi et al., 2014). The human structure-defined RRM of RBPMS spans from residue 20 to 111 corresponding to the N-terminus of the RBP. In brief, RBPMS homodimerization interface is formed by electrostatic and hydrophobic interactions, with Asp34, Lys36, Arg38, Glu39 and Asp87 acting as key residues. The RRM residues involved in establishing the intermolecular interactions are highlighted in Figure 7.1. Apart from the RRM, no further predicted structured domains are found in the RBPMS sequence. In fact, the C-terminus of

RBPMS consists of an intrinsically disordered proline rich region that could potentially lead to further RNA or protein interactions. Moreover, despite its predominant nuclear localization in PAC1 cells, no evident NLS could be identified in RBPMS sequence. Lastly, RBPMS is found expressed as two main isoforms, RBPMS-A and RBPMS-B, as a result of an alternative 3′ end exon. Hence, RBPMS isoforms only differ in their C-terminal tails, in which RBPMS-A codes for a 20 amino acid sequence whereas RBPMS-B contains a slightly longer tail formed by 43 residues.

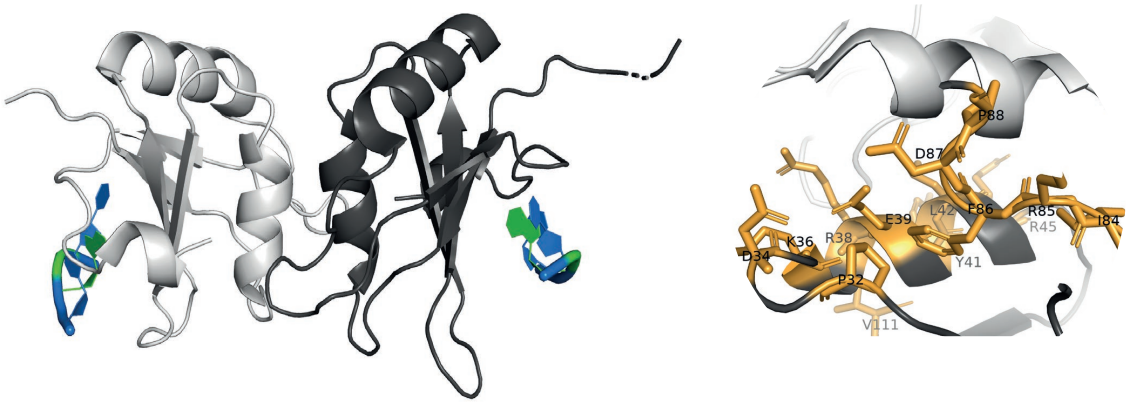


Fig. 7.1 Structure of human RBPMS dimer bound to CAC RNA molecule **Left**, crystal structure of the RBPMS homodimer (gray and black) bound to two CAC RNA molecules (blue and green representing C and A respectively). **Right**, dimerization interface with the residues involved in the dimerization labeled and highlighted in orange. Images were generated using PyMol (v2.3.0) and the deposited structure of RBPMS RRM–RNA complex from the Research Collaboratory for Structural Bioinformatics PDB (accession code 5DET).

Therefore, even though only a single RRM is present in RBPMS, its dimerization confers a modular arrangement typical of other RBPs (Lunde et al., 2007). Consequently, RBPMS dimerization promotes cooperative interaction strengthening the protein-RNA binding affinity as indicated by structural-functional analysis of dimer mutants (Farazi et al., 2014). Dimerisation is also known to allow RBPs to engage in simultaneous protein-protein and protein-RNA interactions (Lunde et al., 2007). Consistent with that, impairment of RBPMS2 dimerization resulted in the loss of the interaction with EEF2 (Sagnol et al., 2014) and reduction of the ESRP1-RBPMS2 interaction (Sagnol et al., 2016).

7.1.2 C terminal disorded domains

Despite the increasing amount of evidence highlighting the importance of RBP's disordered regions (Calabretta and Richard, 2015), structural studies regarding these protein segments are still limited due to the technical challenges imposed by these regions. Thus, although the structure of RBPMS has been solved, the implications of its C-terminus in the function and interactions of the protein remains unclear. Nevertheless, functional studies with C-terminus deleted mutants have revealed its importance to RBPMS functions. In contrast with the structural work, deletion of RBPMS C-terminus (100-197) disrupted both RNA binding and dimerization in *Xenopus* oocytes as indicated by *in vivo* poly(A) binding assays and immunoprecipitation (Gerber et al., 2002). Deletion of the conserved C-terminal domain while maintaining the RRM domain of RBPMS2 also altered its subcellular localization in zebrafish somatic blastula cells (Kaufman et al., 2018) and in *Xenopus* retinal ganglion cells (RGC) (Hörnberg et al., 2013). Moreover, RBPMS C-terminus was required for full stimulation of TGF-β/Smad-regulated transcription and sufficient to interact with cFos in HEK293T cells (Fu et al., 2015; Sun et al., 2006). Therefore, further studies focusing on RBPMS C-terminus could reveal interesting features of the disordered regions in this RBP, e.g. short linear motifs involved in protein-protein interactions.

7.1.3 Post-translational modification in RBPs

Protein post-translational modification (PTM) is a process in which different functional groups are covalently added to the polypeptide sequence, such as phosphate (PO_4^{3-}), methyl (CH_3), acetyl (C_2H_3O), ubiquitin (small protein) and glycans (polysaccharides) (Lovci et al., 2016). The highly dynamic and largely reversible nature of PTMs provides the cell with a mechanism to finely regulate its protein network (Deribe et al., 2010). In terms of function, PTMs can affect translocation, secretion, function and elimination of the modified protein, leading to effects in almost all cellular processes, including survival, proliferation, differentiation and migration (Deribe et al., 2010; Lovci et al., 2016). Interestingly, many RBPs could have their RNA-binding activity regulated by PTMs as indicated by the strong enrichment of modification sites within RBPs (Castello et al., 2016; Hentze et al., 2018). This could consequently affect the dynamics of these RBPs.

PTMs are triggered in response to several extra and intracellular signals in a manner that PTM control of regulatory RBPs allows post-transcriptional processes to respond to stimuli (Hentze et al., 2018; Lovci et al., 2016). For instance, AS response to distinct

extracellular signals like EGF (Epidermal growth factor), Wnt, insulin, cytokines and heat stress act through regulation of splicing factors (Lovci et al., 2016). A splicing regulator affected by PTM via signaling is HNRNPL (Vu et al., 2013). The PI3K/AKT pathway leads to phosphorylation of HNRNPL S52 which then increases its splicing activity by out-competing HNRNPU binding (Vu et al., 2013).

In fact, during VSMC phenotypic switch, cells are exposed to several signal inputs from adhesion receptors, extracellular matrix and growth factors (Frismantiene et al., 2018). Thereby, protein PTM assumes a critical role in the downstream signaling pathway, modulating different targets which include RBPs (Lovci et al., 2016).

Thus, the work described in this chapter aimed at mapping RBPMS regions involved in its splicing activity using structural-functional analysis. Additionally, motivated by the importance of PTMs in the regulation of RBPs, we sought to uncover any potential modulation of RBPMS splicing function via PTMs, such as phosphorylation and ubiquitination.

7.2 Results

7.2.1 RBPMS dimerization is critical for its splicing regulation

In order to understand the contribution of the dimerization to RBPMS splicing function, two RBPMS-A dimer mutants were created by mutation of either individual residue R38 or combined with E39 (R38Q and R38E & E39A, dimer mutant 1 and 2 respectively) (Figure 7.2A). These mutants were previously shown to impair dimerization and localization to stress granules (Farazi et al., 2014). Transfection of HEK293 cells with RBPMS wild-type and dimerization mutants was then carried out to assess any functional consequence to the repression of *Tpm1* minigene and *ACTN1* and activation of *FLNB* and *MPRIP* splicing (Figure 7.2B-D). All proteins were expressed to comparable levels as indicated by western blot, allowing comparison of their activities. Consistent with previous data in this study, RBPMS overexpression in HEK293 cells promoted the SM splicing pattern with RBPMS isoform B showing less activity in the repression events, *Tpm1* and *ACTN1*. Conversely, both mutations affecting dimerization of the most active isoform, RBPMS-A, completely abrogated its splicing activity, both repression and activation as shown by the inclusion levels similar to the basal level observed for the controls and by the splicing efficiency (Figure 7.2 B and D). Splicing efficiency of the mutants was calculated by subtracting the Venus

inclusion level from the inclusion levels of the different RBPMS constructs, then the difference was normalized to the most active isoform (RBPMSA). The mean of the PSI values were used to calculate the splicing efficiency. RBPMS binding mutant (K100E) was transfected alongside, since mutation of residues involved in dimerization were reported to decrease RBPMS RNA binding (Teplova et al., 2016). Both RNA binding and dimer mutants showed similar results. Thus, RBPMS regulation of splicing not only requires binding to the target RNAs but also relies on its dimerization.

7.2.2 RBPMS C terminus is necessary for its splicing function

In order to map the functional regions of RBPMS, several deletion mutants were designed to cover both N and C termini (Figure7.3 A). Activity of the mutants in the regulation of splicing was tested in HEK293 cells. *Tpm1* minigene reporter and endogenous *ACTN1* were used as repressed events and *Myocd* minigene reporter and endogenous *FLNB* as activation ASEs. For repression, only RBPMS ΔN mutant showed activity comparable to wild-type RBPMS-A in both targets, promoting nearly a complete switch in *Tpm1* splicing (Figure7.3 B). Deletion of RBPMS C-terminus, RBPMS Δ20, leaving all sequences in common to RBPMS-A and B, resulted in partial activity as indicated by the inclusion levels observed for *Tpm1* exon 3 and *ACTN1* exon NM, ∼50% and ∼86% respectively. This mutation is also equivalent to deletion of the RBPMS-B 43 amino acid tail, and compared to full length RBPMS-B it has increased activity (79.8% versus 53.2%). This suggests the existence of an inhibitory activity by the RBPMS-B C-terminal tail. Moreover, deletion of just an additional 15 amino acids (Δ35) was sufficient to drive complete loss of repressor activity (Figure7.3B-C). For activation, a similar pattern was observed, in which the N-terminus was dispensable for its activity and the C-terminus critical for RBPMS regulation of *Myocd* and *FLNB* splicing (Figure7.3 B-C). The full C-terminus deletion mutant, RBPMSΔC, showed severe impairment of its activation (∼7.6% of RBPMS-A activity). However, both RBPMSΔ20 and Δ35 did not impair RBPMS activation to the same extent than observed for repression. RBPMSΔ35 still showed ∼60% and ∼91% activation activity on *FLNB* and *Myocd* splicing respectively, whereas the same mutant displayed less than ∼15% repressing activity on *Tpm1* and *ACTN1* splicing compared to RBPMS-A (Figure7.3 C). Finally, RBPMS RRM itself, which is involved in the RNA binding and dimerization of RBPMS, was not sufficient to either activate or repress splicing. Therefore, these data provide insights into the regions that are involved in RBPMS splicing in addition to the RRM, uncovering the importance of the C-terminal sequence.

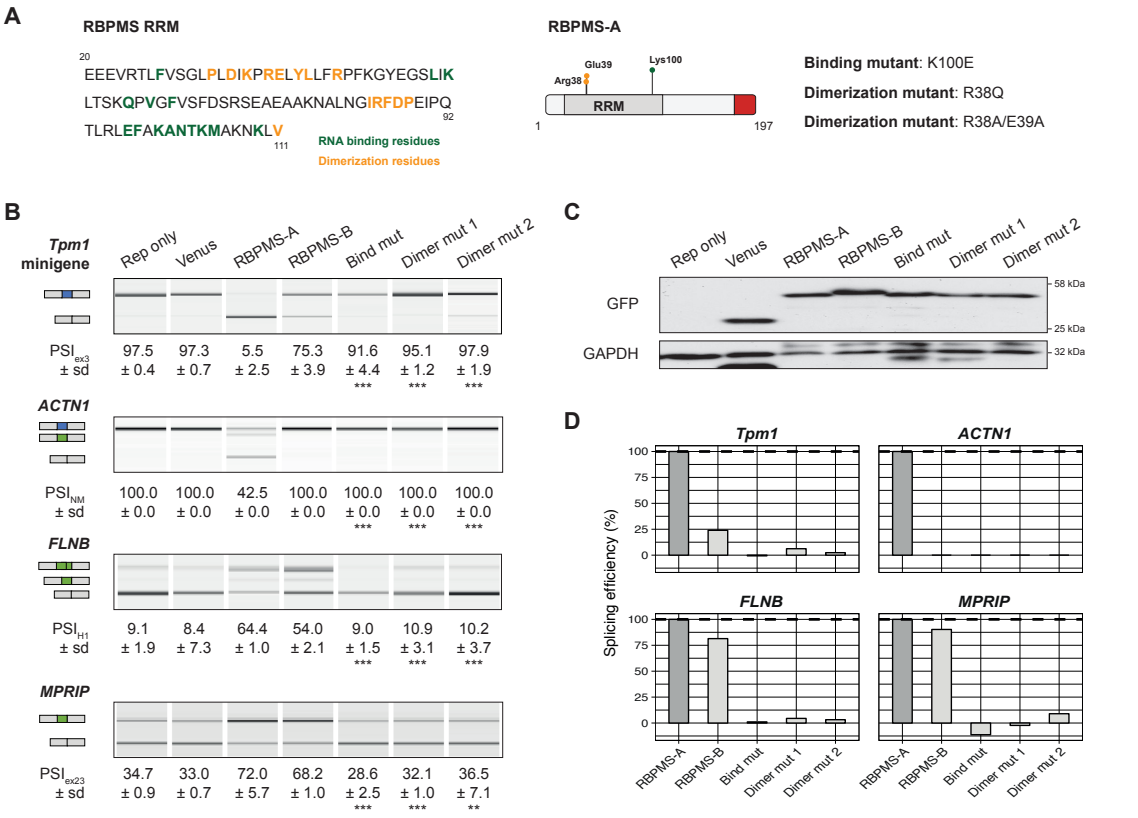


Fig. 7.2 Dimerization is critical for RBPMS-A splicing regulation. A Left, amino acid sequence of RBPMS *R*NA *R*ecognition *M*otif (RRM) with residues involved in RNA binding and dimerization highlighted in green and yellow. Right, full length RBPMS schematic with mutated residues indicated. **B** RT-PCRs to verify the splicing patterns of rat *Tpm1* exon 3 reporter and endogenous *ACTN1*, *FLNB* and *MPRIP* upon overexpression of rat RBPMS-A and RBPMS-B wild-type and binding and dimerization RBPMS-A mutants in HEK293 cells. Reporter only and Venus controls were also included. PCR products are indicated on the left. PSI values shown are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. **C** Western blots probing for GFP and GAPDH as a loading control. Protein size markers are shown on the right. **D** Bar plots of the splicing efficiency of the different effectors normalized to RBPMS-A isoform (dark gray). Means of the PSI values were used to calculate the splicing efficiency.

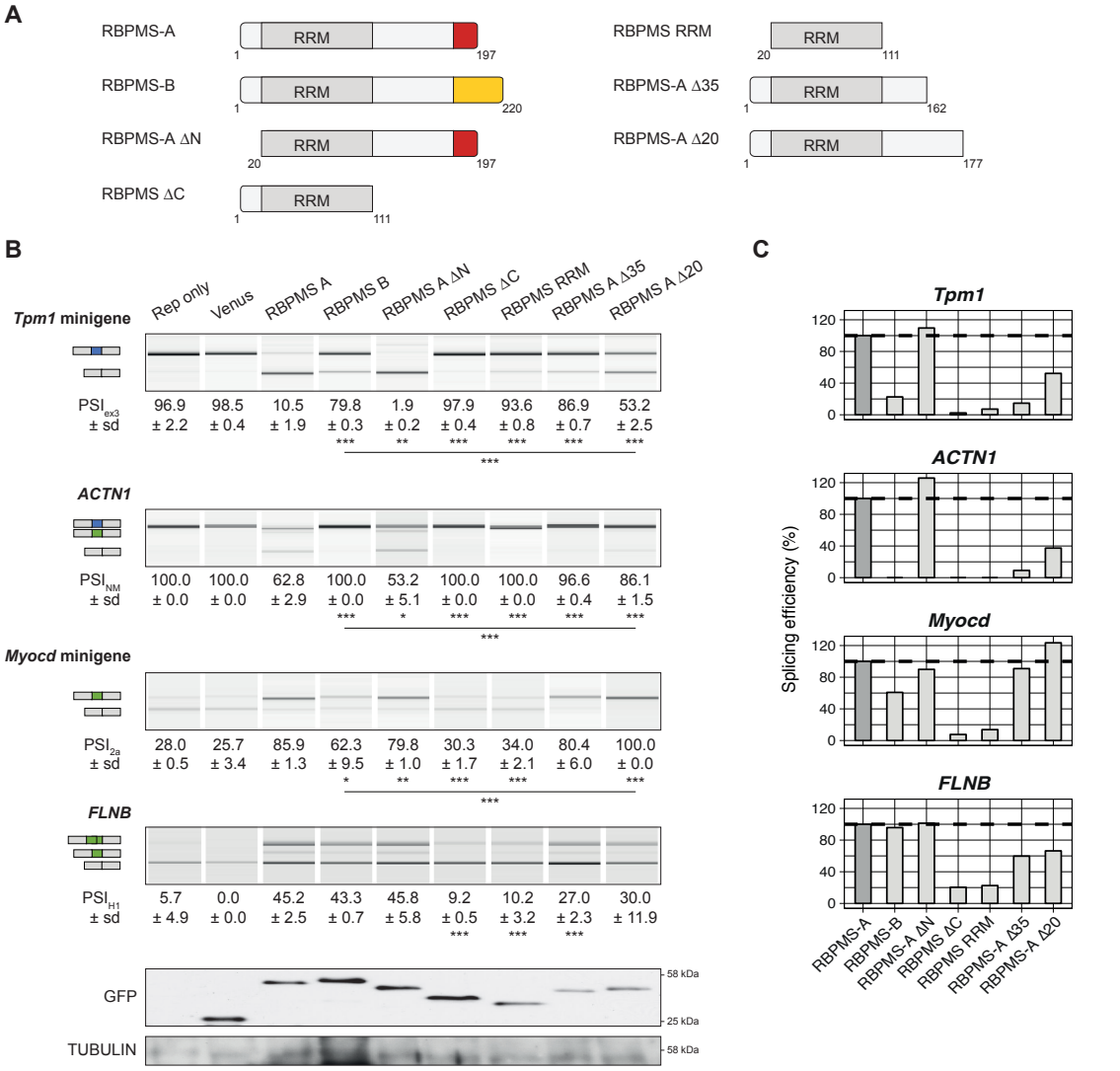


Fig. 7.3 RBPMS C-terminus is required for splicing regulation. A Schematics of the different RBPMS mutants tested. **B** Top, activity of RBPMS mutants upon *Tpm1* and *Myocd* minigene reporters and endogenous *ACTN1* and *FLNB* in HEK293 cells were assessed by RT-PCR. Reporter only and Venus controls were analyzed in parallel as well as wild-type RBPMS isoforms. PCR products are indicated on the left. PSI values are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Bottom, western blots probing for GFP and TUBULIN as a loading control. Protein size markers are shown on the right. **C** Bar plots of the splicing efficiency of the different effectors normalized to RBPMS-A isoform (dark gray bar). Means of the PSI values were used to calculate the splicing efficiency.

In addition to the N and C terminal mutations, Zoe Heckhausen, a summer student under my day to day supervision in the laboratory, obtained RBPMS-A mutants containing internal deletions in regions of RBPMS C-terminus (Figure 7.4). Internal deletion regions were chosen by the level of conservation across different species (Figure 7.4). Residues 126 to 146 were deleted in RBPMS-A Δ1, 148 to 163 from RBPMS-A Δ2 and a larger deletion, 126 to residue 163, in the case of RBPMS-A Δ1-2. Regulation of *Tpm1* exon 3 splicing by RBPMS-A mutants was verified in HEK293 cells transfections. The wild-type RBPMS showed the usual differential activity and despite all the deletions, all RBPMS-A mutants showed no difference in their repressive activity relative to RBPMS-A wt, splicing efficiency of ∼90% (Figure 7.4). Therefore, RBPMS repressive activity is driven by further sequences embedded in its C-terminus than the ones deleted in here.

7.2.3 Assessing RBPMS activity by artificial tethering

To further investigate the contribution of the location of the binding sites to the splicing regulation as well as a starting point for further structure function analysis, artificial MS2 tethering assays were carried out. This assay consists of recruiting RBPs to RNA molecules via the interaction between the MS2 bacteriophage coat protein domain and its specific stem-loop structure motif (Figure 7.5).

Firstly, full length wt RBPMS-A, RBPMS-B and, RBPMS-A binding mutant were fused to the bacteriophage MS2 protein domain. Additionally, MS2-coat protein hairpin binding sites were inserted downstream and upstream of the mCAC *Myocd* and *Tpm1* minigene reporters respectively (Figure 7.6). Restoring of the activation of *Myocd* exon 2a and repression of *Tpm1* exon 3 was then assessed by overexpression of MS2-tagged RBPMS proteins in HEK293 cells (Figure 7.6). RBPMS recruitment to either CAC locations downstream of *Myocd* exon 2a was sufficient to restore its inclusion (∼45%) (Figure 7.6 top panel). Moreover, the RNA binding mutant, K100E, was also able to activate splicing of exon 2a but to a lesser extent (∼20%) (Figure 7.6 top panel). For *Tpm1*, individual hairpins in each cluster upstream of exon 3 resulted in very mild effects, being the greatest repression achieved with the further CAC location, 25.5% and 15.6% exclusion of exon 3 for RBPMS-A and RBPMS-B respectively (Figure 7.6 bottom panel). Insertion of the hairpins in both locations of *Tpm1* caused a greater level of skipping, approximately 50% and 40%, but still not as fully responsive as the wild-type constructs when transfected with RBPMS-A. On the other hand, RBPMS-B showed higher repressive activity than in the previous transfections. Interestingly, recruitment of the binding mutant to the location did not affect the inclusion levels of

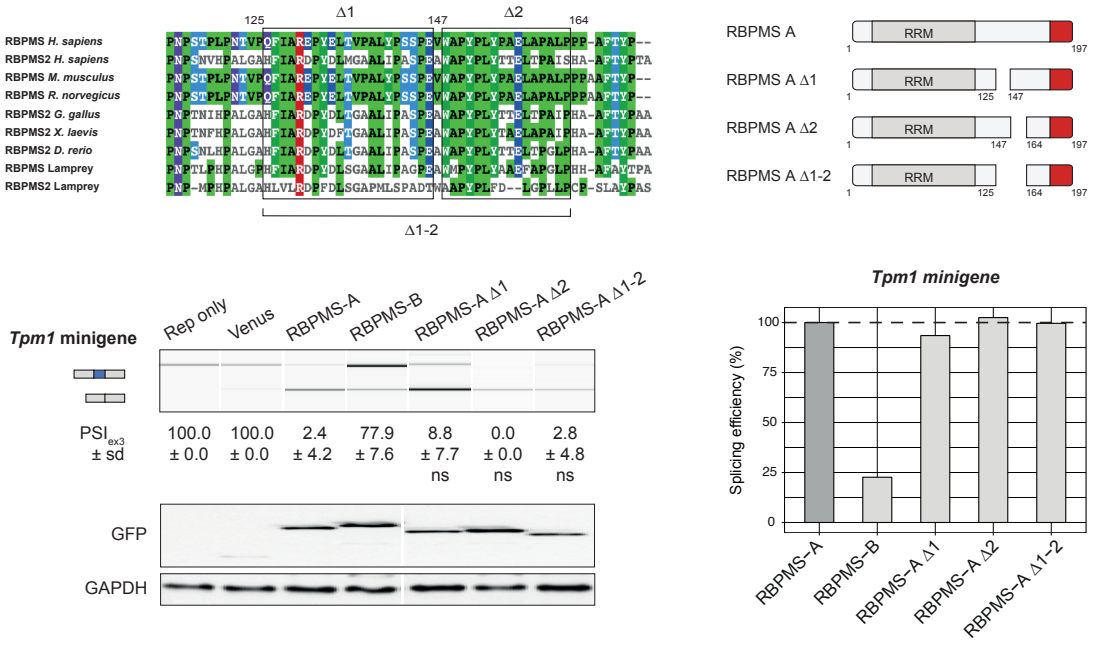


Fig. 7.4 Internal deletions of RBPMS C-terminus do not affect its splicing activity. Top RBPMS alignment and conservation across different species and schematics of RBPMS-A internal deletions. **Bottom** Activity of RBPMS mutants upon *Tpm1* minigene reporter in HEK293 cells were assessed by RT-PCR. Reporter only and Venus controls were analyzed in parallel as well as wild-type RBPMS isoforms. PCR products are indicated on the left. PSI values are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Bottom, western blots probing for GFP and GAPDH as a loading control. Protein size markers are shown on the right. **C** Bar plot of the splicing efficiency of the different effectors normalized to RBPMS-A isoform (dark gray bar). Means of the PSI values were used to calculate the splicing efficiency.

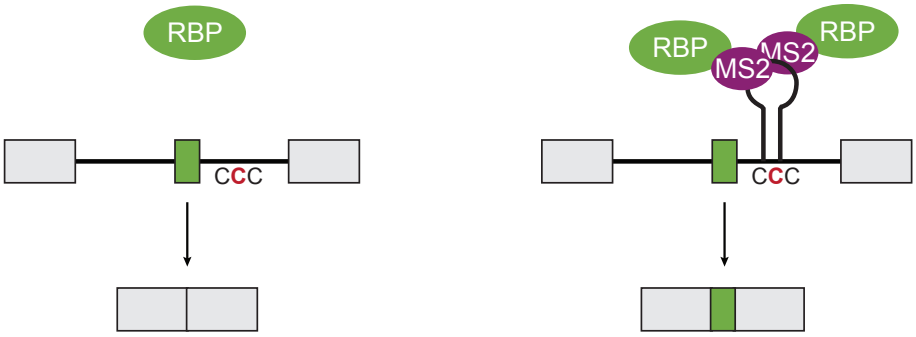


Fig. 7.5 MS2 tethering assay. Tethered function technique allows elucidation of the molecular roles of RBPs by artificial recruitment of these proteins to sites in RNA molecules via the interaction between the bacteriophage MS2 domain and its hairpin-binding motif.

exon 3. This suggests that perhaps the RNA binding capacity from RBPMS RRM is still somehow required for its function, however this was not further investigated herein. Moreover, another explanation could be that more RBPMS needs to be recruited to repress splicing. Indeed *Tpm1* exhibits several CAC motifs upstream of exon 3 (Figure 3.11) and insertion of further MS2 hairpins might be necessary for RBPMS maximal repression. In summary, recruitment by artificial tethering recovered RBPMS activation whereas only partial repression was rescued as indicated by *Myocd* exon 2a inclusion and *Tpm1* exon 3 skipping respectively.

To get more insights into RBPMS structure functions, *Myocd* mCAC minigene reporter containing the MS2 hairpins was then further used to recruit MS2-tagged RBPMS mutants. RBPMS-A and B with N-terminus and RRM deleted (RBPMS-$A_{112\text{-}197}$ and RBPMS-$B_{112\text{-}220}$), RBPMSΔC and, RBPMS RRM were tested upon their regulation of *Myocd* exon 2a inclusion by artificial recruitment (Figure 7.7). Interestingly, the RBPMSΔC mutant and the RRM on its own were not sufficient to promote inclusion of the exon 2a (PSI_{ex2a}= 0.0). Nevertheless, both RBPMS-$A_{112\text{-}197}$ and RBPMS-$B_{112\text{-}220}$ mutants retained more splicing activity than the other RBPMS deletion-mutants (Figure 7.7), despite the lack of the RRM domain and their lower protein expression levels compared to the other mutants (see western blotting). However, their activity was still lesser than the wt proteins and RBPMS binding mutant. Thus, these data further confirmed the requirement of the C-terminal sequences of RBPMS for its mediated splicing. It will be interesting to assess whether this is a common feature of RBPMS repression as well.

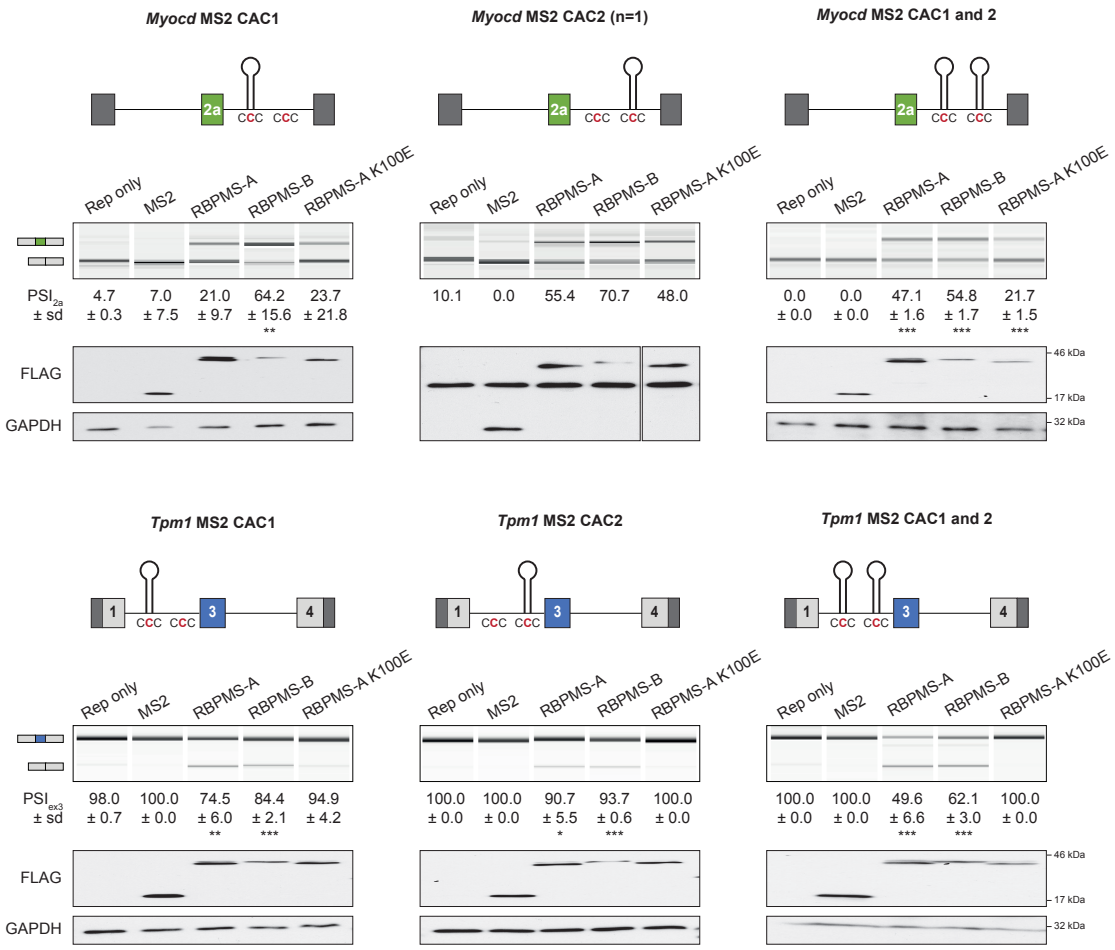


Fig. 7.6 Regulation of mutated reporters are rescued by artificial tethering assays. MS2 hairpins were inserted into the *Myocd* and *Tpm1* minigene reporters as shown in the schematics, top and bottom respectively. Their activities were then tested for RBPMS-A and RBPMS-B and binding mutant (K100E) by RT-PCR. Reporter only and MS2 controls were analyzed in parallel. PCR products are indicated on the left. PSI values shown are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Western blots probing for FLAG and GAPDH as a loading control. Protein size markers are shown on the right.

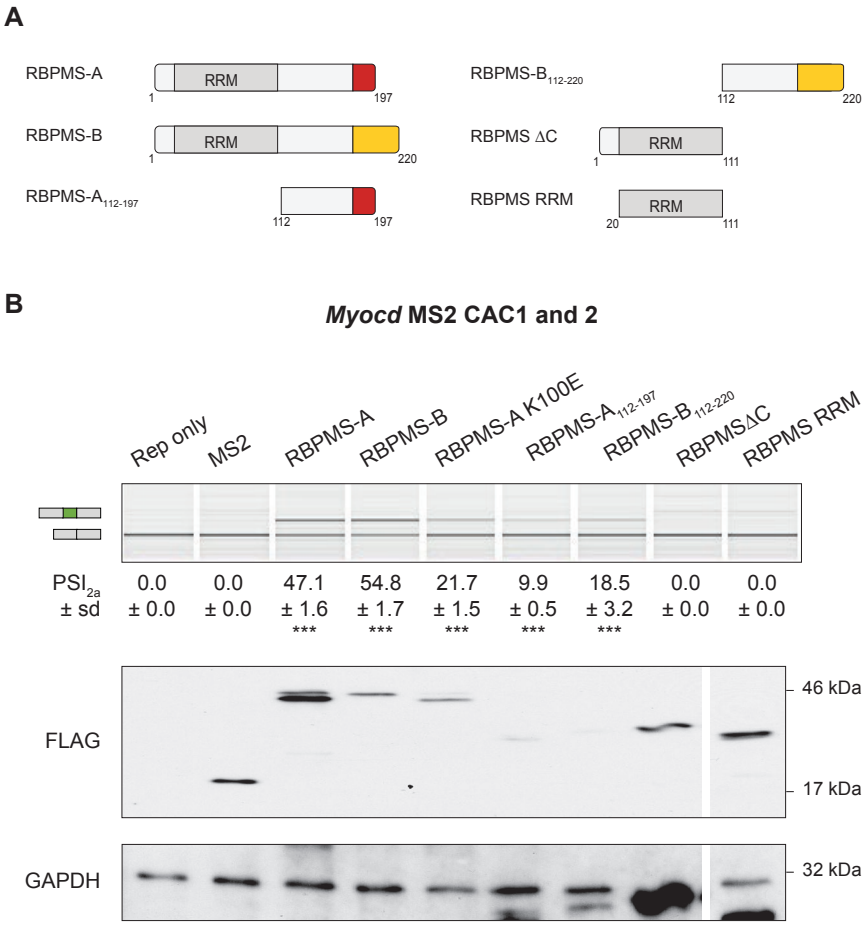


Fig. 7.7 RBPMS regions sufficient for splicing. A Schematics of MS2 tagged RBPMS deletion mutants and wt. **B** Activity of the mutants were then tested alongside RBPMS-A and RBPMS-B and RNA-binding mutant (RBPMS-A K100E) by RT-PCR. Reporter only and MS2 controls were analyzed in parallel. PCR products are indicated on the left. PSI values shown are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Western blots probing for FLAG and GAPDH as a loading control. Protein size markers are shown on the right.

7.2.3.1 Properties of RBPMS isoforms C termini

To explore the importance of the C-terminal tail of each isoform, more mutational functional studies were carried out focusing on RBPMS-A and RBPMS-B last 20 and 43 amino acids. RBPMS-B C-terminal tail was fused to RBPMS-A, creating an RBPMS-AB mutant, and vice versa (Figure 7.8A). These constructs were then co-transfected with *Tpm1* minigene reporter and their splicing activity compared to the wild-type RBPMS isoforms and the C-terminus truncated RBPMS, RBPMS ΔC (Figure 7.8B-C). Interestingly, fusion of the C-terminal tail from the other isoform was sufficient to override the original splicing activity. In other words, RBPMS-A fused with RBPMS-B C-terminus lost its full repression of *Tpm1* exon 3 (PSI= 59.9%), whereas the RBPMS-B containing the RBPMS-A tail was then able to repress splicing of exon 3 (PSI_{ex3}= 12.7%) (Figure 7.8B). Moreover, the C-terminus deletion remained more active than RBPMS-B or AB indicating an inhibitory function of isoform B tail. In summary, this showed that the full repressive capacity of RBPMS is determined by the isoform C-terminal tail; the RBPMS-A tail enhances activity, while the RBPMS-B tail reduces activity of the remainder of the protein.

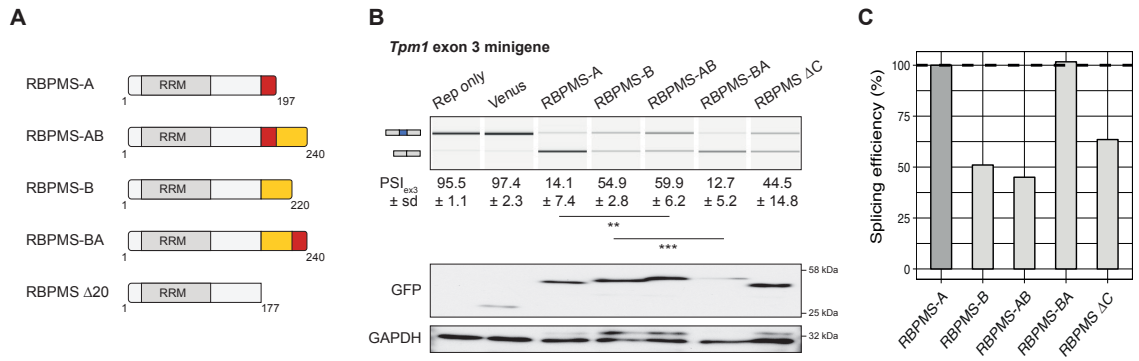


Fig. 7.8 Further C-terminus of RBPMS determines its splicing repression capability. **A** Schematics of the RBPMS mutants. **B** Top, activity of RBPMS mutants upon rat *Tpm1* minigene reporter in HEK293 cells were assessed by RT-PCR. Reporter only and Venus controls were analyzed in parallel as well as wild-type RBPMS isoforms. PCR products are indicated on the left. PSI values are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Bottom, western blots probing for GFP and GAPDH as a loading control. Protein size markers are shown on the right. **C** Bar plots of the splicing efficiency of the different effectors normalized to RBPMS-A isoform (dark gray bar). Means of the PSI values were used to calculate the splicing efficiency.

7.2.3.2 RBPMS-B and its potential inhibitory role

A possible explanation for the RBPMS-B C-terminus-inhibition of its repressive activity is that the C-terminal tails have different properties such as aggregation capabilities. Aggregation of RBPs has been studied in pathologies mainly neurodegenerative diseases (e.g. FUS and TARDBP in amyotrophic lateral sclerosis - ALS), but more recently it was also implicated in the regulation of RBP splicing function, for instance RBFOX self-aggregation through its tyrosine-rich domain (Conlon and Manley, 2017; Ying et al., 2017). So, using a protein aggregation predictor, AGGRESCAN (Conchillo-Solé et al., 2007), the differences in the "aggregation" propensity between the two major isoforms of RBPMS were inspected (Figure 7.9A). Strikingly, RBPMS-B C-terminus, highlighted in yellow, was predicted to be more prone to formation of aggregate structures as indicated by the high HSA (hot spot area) value (Figure 7.9A). On the other hand, this feature was not predicted for RBPMS-A. Thus, to address the importance of the aggregation, the residues predicted to be involved in this property (LLQQIRFVSGNVFVTYQ) were deleted from RBPMS-B C-terminus, RBPMS-B HSA. This mutant was then tested upon regulation of the rat *Tpm1* exon 3 minigene reporter. Remarkably, the mutant RBPMS-B HSA, gained full repressive activity as indicated by the PSI= 34.2% similar to the RBPMS-A PSI= 36.0% whereas RBPMS-B displayed a PSI= 87.7% (Figure 7.9B and D). Therefore, these data show that the core region of the RBPMS-B specific tail antagonizes the intrinsic activity of the protein. Whether this antagonism is mediated by aggregation remains to be determined.

7.2.4 Post-translational modifications

To address any potential role of post-translation modification (PTM) in the regulation of RBPMS splicing activity, RBPMS residues reported to be post-translationally modified were identified using the PhosphoSitePlus database (Table 7.1) (Hornbeck et al., 2015). The HTP (High Throughput Papers) index was then used to sort the more meaningful PTMs, leading to the following modifications: ubiquitination of K109 and phosphorylation of residues T113 and T118 (Table 7.1). These modifications were in residues lying in both RBPMS isoforms, adjacent to the RRM domain, and therefore could potentially modulate both proteins. Furthermore, only ubiquitination of K191 and methylation of R198 were specific to RBPMS isoform A and RBPMS-B respectively (Table 7.1). The role of phosphorylation was first chosen to be further characterized.

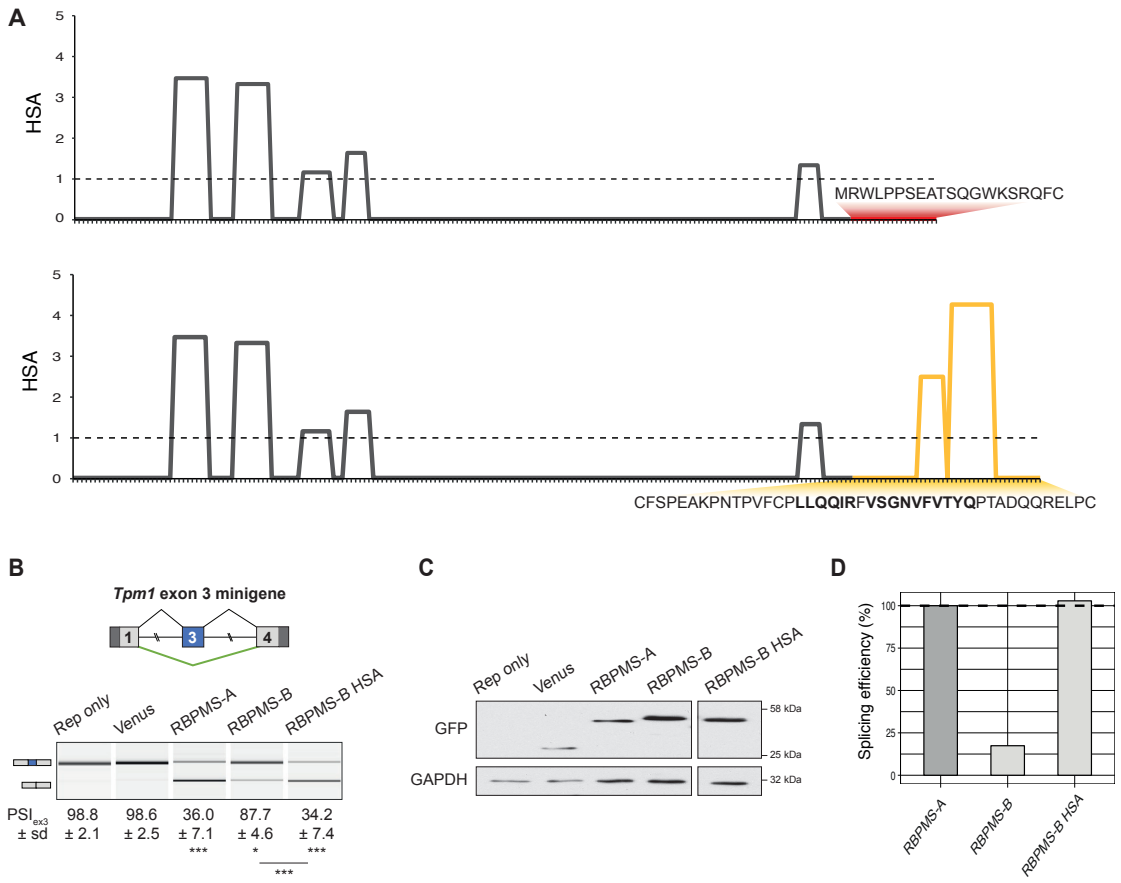


Fig. 7.9 RBPMS isoform specific C-termini have different aggregation propensity. A Plots of the aggregation-prone segments in RBPMS-A and RBPMS-B, top and bottom, respectively. Aggregation prediction is indicated by the "hot spot" area (HSA). Isoform specific C-terminal sequence is highlighted and residues deleted from RBPMS-B are in bold font. Prediction was obtained from AGGRESCAN (http://bioinf.uab.es/aggrescan). **B** Activity of RBPMS-B aggregation mutant upon *Tpm1* minigene reporter in HEK293 cells was assessed by RT-PCR. Reporter only and Venus controls were analyzed in parallel as well as wild-type RBPMS isoforms. PCR products are indicated on the left. PSI values are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. **C** Western blots probing for GFP and GAPDH as a loading control. Protein size markers are shown on the right. **D** Bar plots of the splicing efficiency of the different effectors normalized to RBPMS-A isoform (dark gray bar). Means of the PSI values were used to calculate the splicing efficiency.

Table 7.1 List of human RBPMS PTM from the PhosphoSitePlus database

Residue	Sequence	RBPMS		HTP
		RBPMS-A	RBPMS-B	
K6	MNNGG**K**AEkENtP	✓	✓	1
K9-ub	NNGGKAE**k**ENtPsEA	✓	✓	2
T12-p	GKAEkEN**t**PsEANLQ	✓	✓	2
S14-p	AEkENtP**s**EANLQEE	✓	✓	5
T25-p	LQEEEVR**t**LFVsGLP	✓	✓	1
S29-p	EVRtLFV**s**GLPLDIK	✓	✓	1
K36	sGLPLDI**K**PRELYLL	✓	✓	2
K48	YLLFRPF**K**GYEGSLI	✓	✓	1
K60-ub	SLIKLTS**k**QPVGFVS	✓	✓	6
K109-ub	NTKMAKN**k**LVGtPNP	✓	✓	9
T113-p	AKNkLVG**t**PNPstPL	✓	✓	8
S117-p	LVGtPNP**s**tPLPNTV	✓	✓	4
T118-p	VGtPNPs**t**PLPNTVP	✓	✓	9
S144-p	VPALYPS**s**PEVWAPY	✓	✓	1
R198	CPLLQQI**R**FVSGNVF	gap	✓	5
K191	EATSQGW**K**SRQFC	✓	gap	1

7.2.4.1 RBPMS phosphorylation

Phosphorylation mutant versions of the most active RBPMS isoform, RBPMS-A, were generated by mutating both threonine residues at positions 113 and 118 (Figure 7.10A). Threonine substitutions to alanine, RBPMS T/A, and glutamic acid, RBPMS T/E, were used to mimic unphosphorylated and phosphorylated RBPMS respectively. Mutants were tested upon transient transfection of HEK293 cells and splicing activity verified for *Tpm1* and *Myocd* minigenes as well as endogenous *ACTN1* and *FLNB*, covering repressed and activated targets (Figure 7.10 B-D). Strikingly, the T/E mutation of RBPMS-A severely impaired its repressor activity as indicated by the significant changes in the *Tpm1* and *ACTN1* splicing (Figure 7.10B). RBPMS T/E promoted less repression of the NM exons in *Tpm1* and *ACTN1* (PSI_{ex3}=64.8% and PSI_{NM}= 55.4%). On the other hand, alanine substitutions were associated with an opposite effect, by slightly but significantly enhancing the repression of the same exons (PSI_{ex3}=7.8% and PSI_{NM}= 33.1%). Furthermore, both mutants were expressed at protein levels similar to the wt proteins, allowing attribution of the differential activity solely to the Thr mutations (Figure 7.10 C). Conversely, the same mutations did not impair the activation of FLNB exon H1 and caused minor effects to *Myocd* exon 2a (Figure 7.10 B).

Additionally, RBPMS-A phospho-mutants still retained more activity than RBPMS-B as shown by the parallel transfection of HEK293 cells with this isoform, even though RBPMS-B was more expressed than the other proteins (Figure 7.10 C). These data suggested that phosphorylation at these sites could be critical PTMs, mainly inhibiting the repressive function of RBPMS-A as summarized in the efficiency plots (Figure 7.10 D).

Due to the fact that RBPMS-A phospho-mutations mainly impaired its repressive function, minor or no effects were speculated for RBPMS-B, which shows only full splicing activation. Surprisingly, RBPMS-B phosphomimetic was no longer able to activate splicing of *FLNB* exon H1 (Figure 7.11 A). RBPMS-B T/E showed basal levels of exon H1 inclusion, ∼8%, whereas the T/A mutant kept an activation level similar to the wild-type (48.8%). Consistent with previous data, RBPMS-B had less repressor activity upon *Tpm1* and *ACTN1* splicing and that was further reduced with the substitution of the threonine residue to glutamate in the case of *Tpm1* exon 3 (Figure 7.11 A). Thus, RBPMS isoforms show different effects upon phosphorylation as indicated by these experiments using RBPMS-B phospho-mutants.

To understand the contribution of each of the threonine residues, individual mutants were created: RBPMS-A T113E and T118E (Figure 7.10A). Mutation was also carried out for S117, adjacent to the phosphorylated T118. Splicing of *Tpm1* exon 3 minigene reporter was assessed in HEK293 upon overexpression of the different mutants (Figure 7.12). Mutation of only one of the threonine residues uncovered T113 as the individual site at which phosphomimetic mutation had the greatest effect, by reducing RBPMS-A activity by ∼50% (Figure 7.12B). However, the phosphomimetic mutation at T118 did not affect skipping of exon 3 by RBPMS as shown by the inclusion level similar to the wild-type RBPMS-A (∼23%) (Figure 7.12A). Moreover, none of the single mutations at the threonine sites were able to recapitulate the same loss of activity observed for the double mutant, suggesting that phosphorylation at both residues is required for full inhibition (Figure 7.12 B). Additional mutation at the S117 to either gluatamic acid or aspartate further inhibited the skipping of exon 3, ∼5% compared to RBPMS-A T/E, the difference being only significant for the TST/D mutant (Figure 7.12 A). Thus, the phosphorylation motif in RBPMS involves both threonines at positions 113 and 118 and potentially the serine at position 117 (TxxxST). The sole contribution of serine 117 as well as the effects to RBPMS-B activation are yet to be investigated.

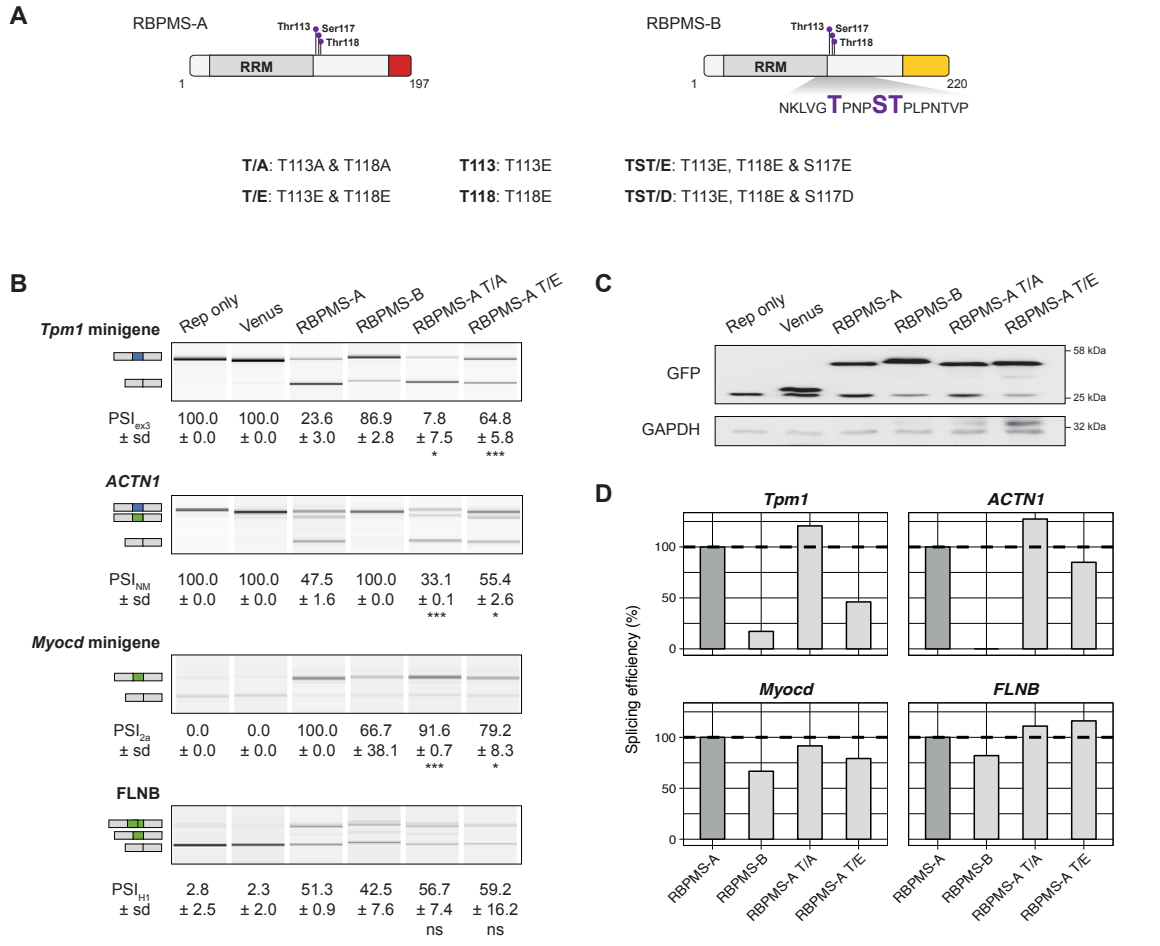


Fig. 7.10 Phosphorylation affects RBPMS splicing activity. A Schematics of RBPMS isoforms A and RBPMS-B with phosphorylated motifs highlighted. Mutants are listed under it. **B** Activities of the mutants upon *Tpm1* and *Myocd* minigene reporter and endogenous *ACTN1* and *FLNB* in HEK293 were assessed by RT-PCR. Reporter only and Venus controls were analyzed in parallel as well as wild-type RBPMS-A and RBPMS-B. PCR products are indicated on the left. PSI values shown are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. **C** Western blots probing for GFP and GAPDH as a loading control. Protein size markers are shown on the right. Lower band in the GFP western blot is a by-product of the *Myocd* minigene reporter. **D** Bar plots of the splicing efficiency of the different effectors normalized to RBPMS-A isoform (dark gray bar). Means of the PSI values were used to calculate the splicing efficiency.

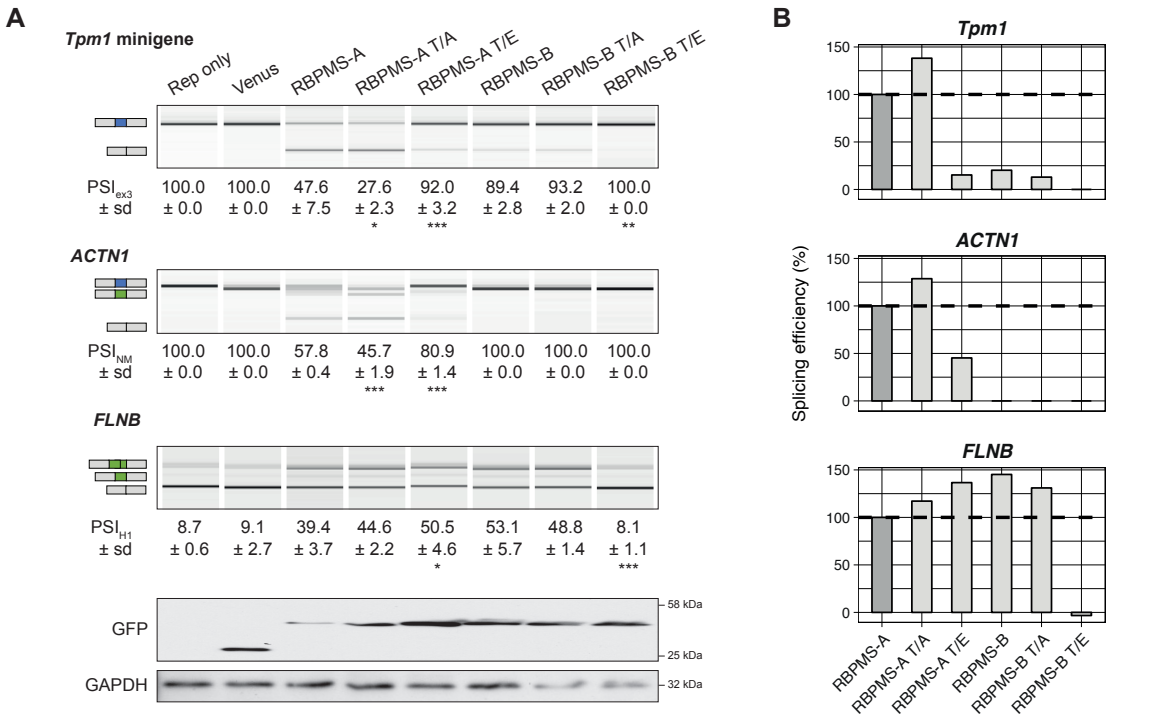


Fig. 7.11 Phosphorylation differentially affects RBPMS isoforms. A Top, activity of RBPMS-B phosphomimetics upon *Tpm1* minigene reporter and endogenous *ACTN1* and *FLNB* in HEK293 were assessed by RT-PCR. Reporter only and Venus controls were analyzed in parallel as well as wild-type RBPMS isoforms. RBPMS-A mutants were also tested alongside. PCR products are indicated on the left. PSI values are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Bottom, western blots probing for GFP and GAPDH as a loading control. Protein size markers are shown on the right. **B** Bar plots of the splicing efficiency of the different effectors normalized to RBPMS-A isoform (dark gray bar). Means of the PSI values were used to calculate the splicing efficiency.

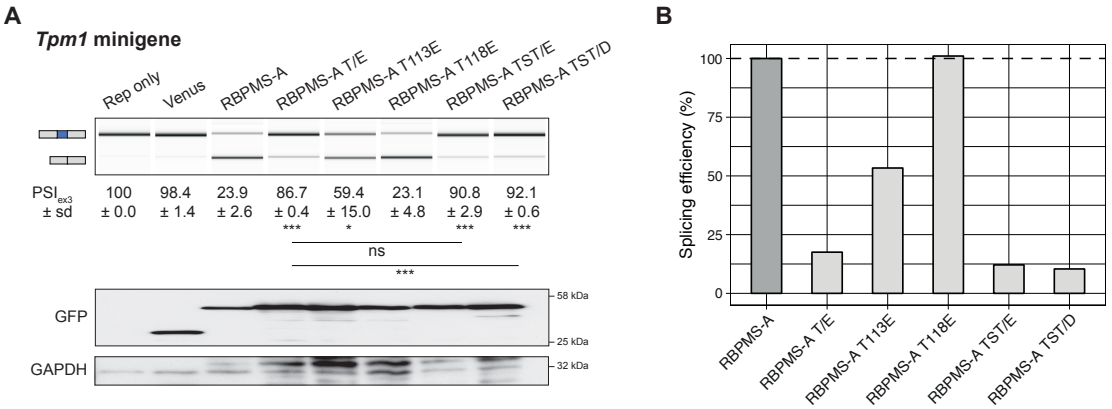


Fig. 7.12 Phosphorylation of both threonines, T113 and T118, is required for maximum inhibition A Top, activity of RBPMS-A single mutants and serine phosphomimetics upon *Tpm1* minigene reporter in HEK293 cells were assessed by RT-PCR. Reporter only and Venus controls were analyzed in parallel as well as wild-type RBPMS-A. RBPMS-A T113E & T118E was also tested alongside. PCR products are indicated on the left. PSI values are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Bottom, western blots probing for GFP and GAPDH as a loading control. Protein size markers are shown on the right. **B** Bar plots of the splicing efficiency of the different effectors normalized to RBPMS-A isoform (dark gray bar). Means of the PSI values were used to calculate the splicing efficiency.

7.2.4.2 RBPMS ubiquitination

Another PTM strongly supported by experimental evidence was the ubiquitination of the lysine 109, just upstream of the phosphorylated residues (Table 7.1). This residue lies within the RRM domain and is involved in contacting RNA (Figure 7.2) (Farazi et al., 2014). The proximity to the phosphorylation sites led to the hypothesis that perhaps its ubiquitination is linked to the initial phosphorylation of these residues in a similar manner to reported for IkappaB (DiDonato et al., 1996). Phosphorylation of IkappaB at two serine residues triggers ubquitination of adjacent lysines that then target it for proteasome degradation (DiDonato et al., 1996). To investigate a similar regulation of RBPMS, lysine 109 to arginine amino acid substitution was carried out in both RBPMS isoforms, wild-types and phosphomimetics (Figure 7.13A). From all the RBPMS ubiquitination mutants, only the RBPMS-A K109R mutant showed significant change in the inclusion level of exon 3 compared to wild-type, being exon 3 more skipped (88.3% exclusion) (Figure 7.13B-C). In addition to that, no increase in the protein expression levels was observed for the ubiquitination mutants (Figure 7.13B, bottom). Actually, if there is any apparent change in expression, it is towards a decrease upon phosphorylation and blockage of ubiquitination as indicated by the levels of RBPMS KT mutants in the western blot (Figure 7.13B, bottom). Therefore this experiment did not provide any evidence supporting a linkage between RBPMS ubiquitination and its protein turnover. Furthermore, the K109R mutation did not block the effect of the T/E mutations.

7.3 Discussion

7.3.1 RBPMS dimerization and its splicing activity

Consistent with studies associating RBPMS dimerization to its functions, disruption of RBPMS dimerization strongly affected its splicing function (Figure 7.6). The same mutant tested in this study was no longer able to localize to stress granules upon oxidative stress (Farazi et al., 2014). Moreover, impairing RBPMS2 dimerization also had deleterious effects to SMC dedifferentiation and interaction with ESRP2 (Sagnol et al., 2016, 2014). Thus, dimerization might be involved in RBPMS functions by also determining protein-protein interactions. Nevertheless, the primary cause of the reported effects is most likely to be the reduction of RNA binding. Dimerization allows effective cooperative binding to dual CACs, but the impairment of the dimer interface

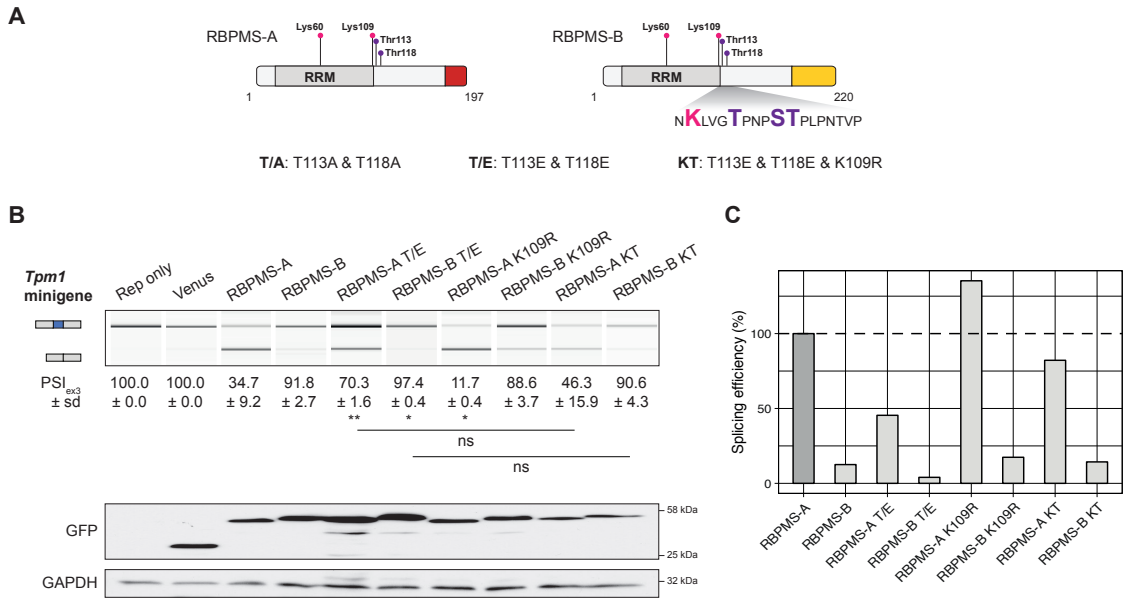


Fig. 7.13 RBPMS phosphorylation is not linked to ubiquitination. A Schematics of RBPMS isoforms and location of lysine residues shown to be ubiquitinated (pink). Mutants are listed below schematics. **B** Top, activity of RBPMS ubiquitination resistant mutants upon *Tpm1* minigene reporter in HEK293 cells were assessed by RT-PCR. Reporter only and Venus controls were analyzed in parallel as well as wild-type RBPMS isoforms. RBPMS-A and RBPMS-B T113E & T118E were also tested alongside. PCR products are indicated on the left. PSI values are mean ± sd (n= 3). Statistical significance was verified by Student's t test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Bottom, western blots probing for GFP and GAPDH as a loading control. Protein size markers are shown on the right. **C** Bar plots of the splicing efficiency of the different effectors normalized to RBPMS-A isoform (dark gray bar). Means of the PSI values were used to calculate the splicing efficiency.

leads to significant decrease in RNA binding affinity. Indeed, Teplova et al. (2016) showed a 4-fold drop in the affinity of dimer mutants.

7.3.2 RBPMS activation and repression mechanisms

Artificial tethering assays of RBPMS using MS2 coat protein further confirmed RBPMS positional-dependent splicing activity and also suggested RBPMS-mediated repression of *Tpm1* exon 3 to be mechanistically distinct from *Myocd* exon 2a activation. Upstream binding of RBPMS was only able to repress splicing in the presence of a functional wt RRM domain (Figure 7.10). At first sight this appear to be a surprising result, since the basis of the MS2 approach is that you can replace the normal mode of RNA binding by an artificial RNA-RBP interaction. Nonetheless, such difference between splicing activation and repression upon tethering RBPs to downstream and upstream sites has also been reported for RBFOX1 proteins (Sun et al., 2012; Ying et al., 2017). Despite that, the exact mechanism for requirement of RNA binding for repression by this RBP still has to be elucidated. It is possible that RBPMS needs to interact with other RNA sequences in addition to the MS2 hairpin inserted in the minigene, in order to establish repressive activity. A possible scenario is the requirement of further binding of RBPMS on the pre-mRNA itself to loop out the exon to be excluded. An even more interesting scenario would be one where RBPMS interacts with snRNAs in the spliceosome. Indeed, RBPMS and U1A proteins share similarities in their intermolecular recognition of CAC element in RNAs (Teplova et al., 2016). U1A is a component of the U1 snRNP, which is involved in pre-mRNA splicing, and interacts with the stem-loop II of U1 snRNA (Howe et al., 1994; Oubridge et al., 1994). U1 snRNA is critical in the recognition of the 5′ splice site during early steps of the spliceosome assembly (Plaschka et al., 2019). In fact, accessory RBPs have been reported to facilitate U1 snRNP recognition of the 5′ splice site promoting exon inclusion (Roca et al., 2013). Additionally, PTBP1 binding to the CU-rich elements flanking the *c-src* N1 exon allows interaction with U1 snRNA, more specifically with stem-loop IV, and this results in exon repression (Sharma et al., 2011). Inhibition of splicing by PTBP1 was then shown to result from formation of a repressed complex that cannot progress to an exon definition complex (Wongpalee et al., 2016). Therefore, RBPMS could exhibit a similar mechanism of repression by competing with U1A and disrupting interactions in the U1 snRNP bound at the regulated 5′ splice site preventing formation of functional spliceosome. Consistent with PTBP1 binding to elements flanking repressed exon, a few conserved CAC motifs are also found downstream of *Tpm1* exon 3 (Figure 3.11), whose importance to RBPMS splicing is yet to be determined.

7.3.3 RBPMS splicing activity beyond its RRM domain

Further structural-functional analysis disclosed RBPMS C-terminus as essential for its splicing activity and MS2 tethering assays hinted that the C-terminus is potentially sufficient to drive its splicing activation. However, stronger conclusion could not be drawn as a result of problems with protein expression of the deletion mutants. In addition to that, despite being sufficient for RNA binding and dimerization *in vitro* (Sagnol et al., 2014; Teplova et al., 2016), the RRM domain alone could not mediate AS changes. Nevertheless, previous studies have shown that the oligomeric state of RBPMS-A and RNA binding affinity are also dependent on its C-terminal region (Farazi et al., 2014). In agreement with its critical role in splicing, the C-terminal region downstream of the RRM has been implicated in several functions of RBPMS/RBPMS2. Granular localization in retinal ganglion cells (Hörnberg et al., 2013) and cFos interaction in HEK293 cells (Fu et al., 2015). Interestingly, *Xenopus* RBPMS2 C-terminus, last 34 amino acids, was required for binding to *Nanos1* RNA *in vivo*. Therefore future work will address whether loss of splicing by C-terminus truncated RBPMS is due to a decrease in RNA binding, dimerization and/or to intrinsic functions of this region.

Like other splicing regulators that display isoform differential activity (Partridge and Carter, 2017; Tabaglio et al., 2018; Wollerton et al., 2001), RBPMS-A strongly activated and repressed AS whereas RBPMS-B acted mainly as an activator. Surprisingly, both isoforms were able to bind to activated and repressed RNAs with similar affinities (Figure 5.11 and 5.12). Therefore, considering the fact that RBPMS isoforms differ only in their C-terminal sequence, a potential explanation for RBPMS-A repressive role is interactions mediated by the isoform specific region. However, RBPMS Δ20 was still able to cause repression ($\sim$40%) suggesting that it is actually RBPMS-B C-terminus that inhibits its splicing repressor role. Consistent with it, fusion of RBPMS isoforms with each others C-terminal regions reproduced the repressive splicing effects of the fused isoform C-terminus. This indicates that the further C-terminal regions can potentially modulate RBPMS activity by enhancing or diminishing its repression. Function conferred by disordered regions of an RBP is not an exclusive feature of RBPMS and it has been described for several other splicing regulators (Calabretta and Richard, 2015). For instance, RBFOX2 C-terminal domain was shown to recruit LASR (Large Assembly of Splicing Regulators) and to be required for splicing (Ying et al., 2017). This RBFOX2 study also revealed another property of this protein which is the formation of higher order assemblies via its tyrosine-rich C-terminal domain (CTD) (Ying et al., 2017). Actually, AS within regions encoding intrinsically disordered segments of splicing regulatory proteins alters their own splicing regulatory activity

and seems to be a recurring mechanism determining splicing regulation (Gueroussov et al., 2017). Indeed, RBPMS isoforms showed distinct aggregation-prone peptides as predicted by AGGRESCAN (Figure 7.9). Deletion of the peptides predicted as potential hot-spots of amyloidal aggregation from RBPMS-B CTD conferred repressive effects not observed with the wild-type RBPMS-B. Thereby, AS control of RBPMS CTD could affect its splicing regulation through formation of multivalent assemblies. The implication of these residues in RBPMS-B differential localization and formation of intracellular inclusions *in vivo* remain to be addressed. Nevertheless, in preliminary biophysical studies using recombinant protein, RBPMS was found to produce large complexes (data not shown). Further work aiming to address the role of aggregation *in vivo* and *in vitro* in RBPMS splicing will shed some light on its mechanism.

7.3.4 Regulation of RBPMS activity by phosphorylation

Master splicing regulators are likely to be tightly regulated by transcriptional, translational and post-translational signals in order to ensure splicing network only in response to the right environmental cues (Jangi and Sharp, 2014). Considering the importance of PTM to modulation of splicing networks by RBPs (Hentze et al., 2018; Jangi and Sharp, 2014), potential PTM of RBPMS was also examined. Using RBPMS with mutantions in Threonine 113 and 118, a role of phosphorylation in the regulation of RBPMS splicing activity was disclosed in addition to its transcriptional control. PTMs are known to modulate RBPs activity by altering their ability to bind target RNAs or proteins as well as their subcellular localization (Naro and Sette, 2013). Actually, the threonine residues found to be phosphorylated are located adjacent to RBPMS RRM and can possibly affect RNA binding or dimerization. However, splicing activation by RBPMS-A was not affected in the phosphomimetic mutant indicating that RNA binding, dimerization and localization are unlikely to be impaired by phosphorylation. Furthermore, PTMs have been pointed out as one of the factors in the control of the phase separation and aggregation behavior of RBPs (Hofweber and Dormann, 2018). Thereby, an intriguing connection between phosphorylation and aggregation is likely to exist and can perhaps explain the differential splicing effects observed with the phosphomimetic substitutions in RBPMS-A and RBPMS-B.

The kinase involved in RBPMS phosphorylation and the signal that triggers it remain to be identified. Although kinase predictors suggested CDK4 as a potential kinase in the phosphorylation of RBPMS, preliminary data using inhibitors did not affect RBPMS-mediated AS (data not shown). Other possible kinases include ERK and p38, members of the conserved signaling pathways that activate the mitogen-

activated protein kinases (MAPKs) (Roux and Blenis, 2004; Venigalla and Turner, 2012). Interestingly, RBPs have been suggested as a point of convergence of MAPK pathways (Venigalla and Turner, 2012). Lastly, the glycogen synthase kinase-3, GSK3, has recently been shown to phosphorylate multiple splicing factors, leading to indirect changes in AS (Shinde et al., 2017). In fact, the GSK3 consensus motif, (pS/pT)XXX(S/T), resembles RBPMS phosphorylation site (TPNPST). Thus it could potentially be the kinase modulating RBPMS activity via its phosphorylation (Shinde et al., 2017). Therefore, further studies are required to address the pathways driving RBPMS phosphorylation in SMC and other cell types.

Finally, similar phosphorylation taking place in RBPMS2 has not yet been reported (Hornbeck et al., 2015). In the paralog, despite the conservation of all the residues, the Thr118 and Ser117 are inverted and the proline adjacent to the site is missing (TPNP**TS**V). Moreover, in human RBPMS2 the second threonine is not conserved. Thus, RBPMS phosphorylation could correspond to a specific mechanism of its regulation that is not shared between paralogs.

Taken together, the data in this chapter uncovered several features of RBPMS regulation of splicing. However, further studies focusing on RBPMS C-terminal tails are required to better understand the isoform-specific differential activity and the importance of each isoform to the cell biology. Additionally, investigations of the mechanism behind RBPMS phosphorylation will be crucial for comprehending how RBPMS is regulated in response to external signals and its following disruptions to RBPMS splicing network.

7.4 Final conclusions

In summary, this chapter addressed some of the mechanisms underlying RBPMS regulation of splicing. It provided insights into the importance of RBPMS dimerization and its C-terminal region. Modulation of RBPMS by post-translational modifications was also investigated, indicating a role for RBPMS phosphorylation. Lastly, RBPMS-B CTD inhibition of its splicing repression could be associated with its potential aggregation capacity. Therefore, these data pave the way for a better understanding of how this SMC master regulator controls splicing and could potentially respond to different cues from the environment.

In conclusion, the main findings of this chapter are:

1. RBPMS dimerization is required for its splicing activity.

2. RBPMS C-terminus is essential for its activity, with RBPMS-A specific C-terminus being necessary for maximal repression, whereas the RBPMS-B C-terminus antagonizes the repressor activity.

3. Artificial recruitment of RBPMS to downstream and upstream sites of CAC mutated *Myocd* and *Tpm1* recovered splicing activation and repression.

4. MS2 tethered RBPMS RNA binding mutant could activate *Myocd* exon 2a but not repress splicing of *Tpm1* exon 3, suggesting that RNA binding is required for repression.

5. MS2 tethered RBPMS-A and RBPMS-B C-termini hinted a potential splicing function by these sequences.

6. RBPMS-B specific aggregation features might explain its reduced splicing repressor activity.

7. Two Thr residues of RBPMS adjacent to its RRM domain can be phosphoryated.

8. T113 and T118 residues phosphomimetic mutants differentially modulate RBPMS activity, inhibiting RBPMS-A splicing repression and RBPMS-B splicing activation.

9. Phosphorylation of both Thr residues is required for maximum inhibition of RBPMS repression.

10. Effects of T113/118 phosphorylation are unlikely to involve subsequent ubiquitination of K109.

In addition to that, this chapter also led to other questions listed below:

- Why is RBPMS binding mutant tethered to MS2 protein not sufficient for splicing repression?
- What is the kinase responsible for RBPMS phosphorylation?
- What is the signaling pathway that triggers RBPMS phosphorylation?
- How does phosphorylation inhibit RBPMS splicing activity?
- What is the mechanism behind RBPMS regulation of splicing?
- How does RBPMS-B C-terminus inhibit its repressive role?

- Do RBPMS proteins form high molecular weight complexes?
- Do RBPMS isoforms display differential aggregation capacities?
- Can RBPMS C-terminus propensity to aggregate explain its lower splicing activity?

The next chapter explore some of these questions as potential future studies of RBPMS.

Chapter 8

RBPMS: a master splicing regulator of VSMC

8.1 General Discussion

8.1.1 Do super-enhancers point the way to tissue-specific regulators?

By using the approach suggested by Jangi and Sharp (2014), we focused on RBPs whose expression is driven by super-enhancers in smooth-muscle tissues. In this manner, RBPMS was identified as a critical splicing regulator of the VSMC AS program. Therefore, this work vindicates Jangi and Sharp's approach for identification of tissue-specific splicing master regulators paving the way for the identification of further regulators in other tissues. The usefulness of the super-enhancer approach is highlighted by the mRNA-Seq data of rat aorta tissue and PAC1 dedifferentiation. Using mRNA-Seq data only, 29 RBPs are found regulated during the phenotypic switch in PAC1 cells (Figure 8.1) and an even larger number in aorta tissue 49 RBPs (Figure 8.2). On the other hand, by focusing on super-enhancers associated RBPs, only nine candidates were suggested as potential regulators of the SMC program (Chapter 3). Additionally, with RBPMS being among so many other regulated RBPs, perhaps it would not have been chosen as the first candidate to be characterized if the super-enhancer approach had not been used. Even though RBPMS could have been identified by mRNA-Seq data, the method applied in this study then allowed investigating a smaller subset of RBPs, facilitating the identification of RBPMS. Additionally, when we started this study, we did not have mRNA-Seq data for rat aorta and/or PAC1 dedifferentiation. In that way this approach may help in situations in which mRNA-

Seq data is not available for the system of interest. Nevertheless, with the recent growing availability of mRNA-Seq data for a range of tissues, a more global and systematic bioinformatic approach could perhaps be used to identify other possible tissue AS master regulators. For instance, using super-enhancer database (Hnisz et al., 2013; Khan and Zhang, 2016) coupled with GTEX RNA-Seq data (Lonsdale et al., 2013), and an RNA-binding protein database (Ray et al., 2009), it may be possible to find whether motifs enriched around different tissue-specific regulated exon correspond to binding motifs of RBPs associated with super-enhancers in the same tissue. Thus, integrating those datasets via computational analysis may provide a more comprehensive understanding of the specific post-transcriptional regulatory components in distinct physiological contexts.

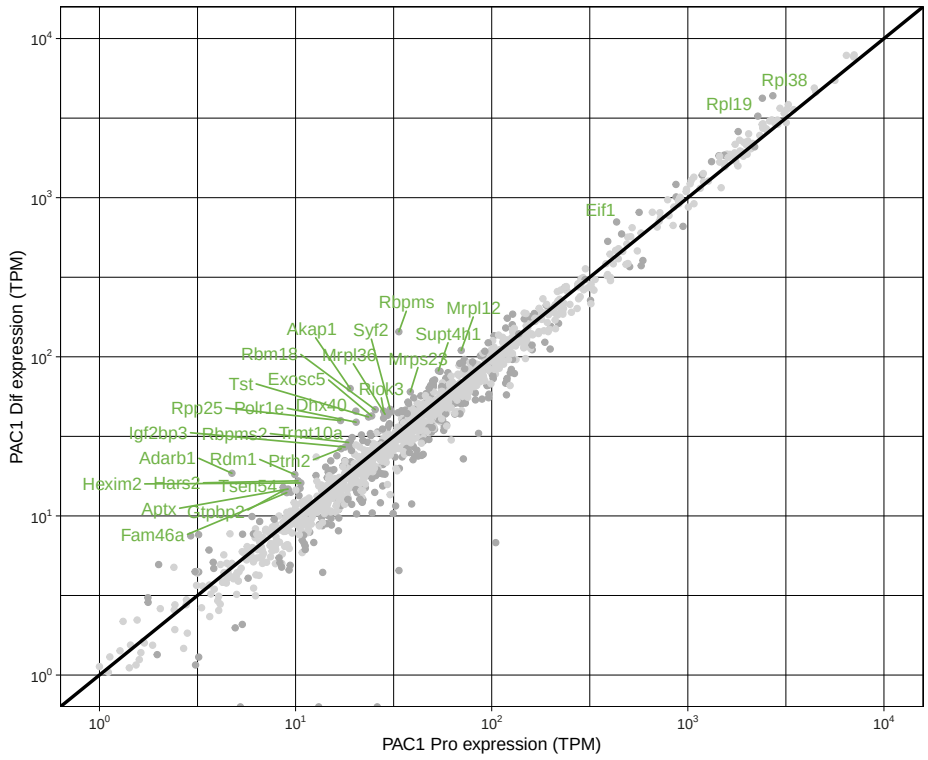


Fig. 8.1 RBPs upregulated in rat differentiated PAC1 cells. RBPs upregulated in rat differentiated PAC1 cells (Dif) compared to the proliferative PAC1 cells (Pro). RBPs showing fold change > 1.5, TPM > 10 and padj < 0.05 are indicated with green labels.

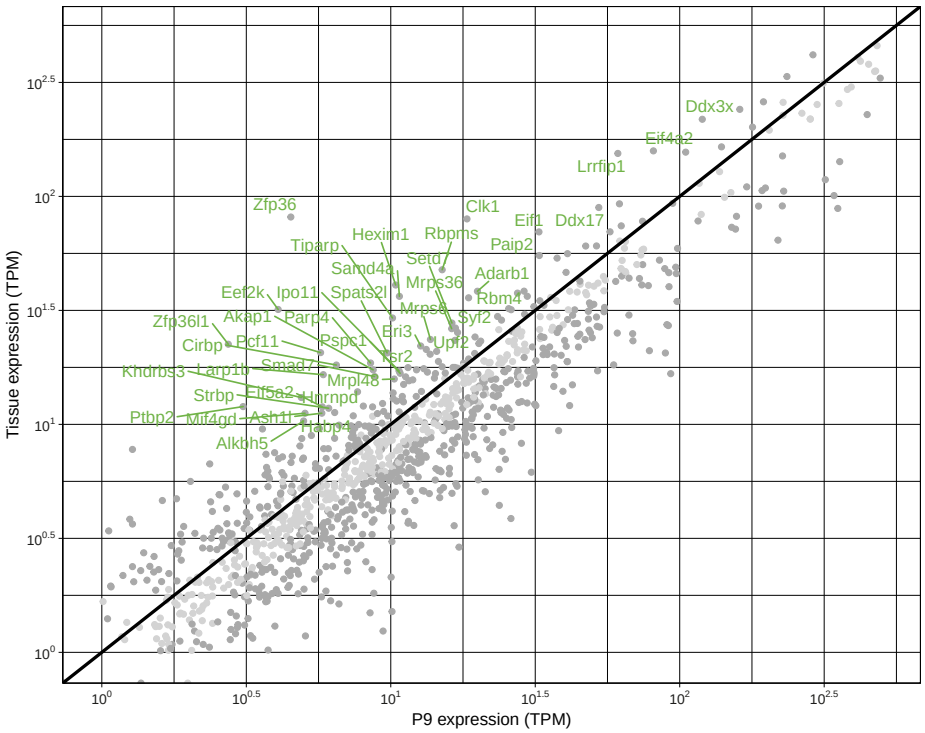


Fig. 8.2 RBPs upregulated in rat aorta tissue. RBPs upregulated in rat differentiated aorta tissue compared to the cultured proliferative VSMCs (T vs P9). RBPs showing fold change > 1.5, TPM > 10 and padj < 0.05 are indicated with green labels.

8.1.2 How much of a master regulator is RBPMS?

RBPMS met many, but not all of the criteria of a master regulator suggested by Jangi and Sharp (2014):

1. it has a wide dynamic range of expression between tissues and differentiation states of SMCs (**Chapter 3**).

2. changes in RBPMS activity appear to be solely responsible for 20% of the AS changes between differentiated and proliferative PAC1 cells (**Chapter 4**);

3. RBPMS target splicing events that are enriched in functionally coherent groups of genes affecting cell-substrate adhesion and the actin cytoskeleton, which are important for SMC cell phenotype-specific function (**Chapter 3-4**);

4. it regulates splicing and activity of other post-transcriptional regulators in SMCs (**Chapter 6**), and

5. it regulates splicing and activity of the key SMC transcription factor MYOCD (**Chapter 6**).

One of the first criteria suggested by Jangi and Sharp (2014) is that master regulators play a role in establishing and maintaining cell identity. However, the biological function of the RBPMS-regulated splicing network to the phenotype of SMCs remains to be addressed. As a splicing regulator, RBPMS is only able to regulate transcripts actively transcribed, and so it is unlikely to be sufficient to establish a differentiated SMC phenotype. Nevertheless, consistent with a role of RBPMS in maintaining a mature SMC differentiated state, RBPMS promoted tissue-like AS patterns, for example in *Cald*, *Fermt2* and *Tns1* (Chapter 4). In fact, the CALD1 isoform produced by RBPMS is a well known marker of SMCs. Thus, even though RBPMS was not enough to upregulate the total transcript levels of SMC markers, it was able to increase SMC-specific isoforms detected at protein levels (Chapter 4). Moreover, RBPMS was responsible for the regulation of a quarter of the AS changes taking place during PAC1 cell dedifferentiation. This is a very impressive number of ASEs solely dependent on RBPMS expression, yet the minimum fraction of splicing events that has to be regulated in a particular condition to qualify the RBP as a master regulator is disputable (Jangi and Sharp, 2014). This quantitative feature could be highlighted in the future by such bioinformatic studies suggested above.

Another feature that RBPMS failed to meet was the lack of AS-NMD events (Chapter 3). In fact, AS regulation of RBPMS can result in non-productive isoforms

that then control its expression levels. Possibly, the presence of regulatory exons is critical in terminally differentiated cells but not in cell-types that retain phenotypic plasticity such as SMCs. The ability of these cells to interconvert between a contractile-differentiated state and a proliferative-dedifferentiated state is required for their physiological functions (Frismantiene et al., 2018). Therefore, perhaps in SMCs these AS-NMD events may consist of the motifs of a bistable double-negative regulation in order to allow master regulators of different phenotypes. Interestingly, QKI which is a regulator of the proliferative phenotype of SMCs might be able to regulate RBPMS expression by affecting production of non-functional and/or NMD targeted transcripts as indicated by the presence of QKI motifs around RBPMS critical exons.

Thus, even though RBPMS did not completely fulfill the criteria of a master regulator as determined by Jangi and Sharp (2014), the features identified in this work provide enough evidence for a role of RBPMS as a major splicing regulator of VSMCs. Future studies will aim to address the remaining characteristics of a master regulator not yet uncovered for RBPMS in this work.

8.1.3 Are changes promoted by RBPMS-A overexpression artefactual?

Insights about RBPMS regulation of SMC splicing were gained by the RBPMS-A overexpression dataset (Chapter 4), but a lot of the overexpression targets are likely to be artefactual. RBPMS-A overexpression led to a strong 70-fold increase in transcript levels and RBPMS-A absolute levels were also incompatible with physiological levels observed in aorta tissue (Chapter 4). Thus, despite being very informative, conclusions drawn from the overexpression dataset require caution. For instance, a lot of genes regulated at the mRNA abundance or the skewed number of ASEs towards less inclusion by RBPMS overexpression were not observed upon RBPMS knockdown. Therefore, due to the fact that the knockdown approach consists of a more physiological approach, it nicely complements the overexpression experiment and reinforces the real effects of manipulation of RBPMS levels.

8.1.4 Does RBPMS phosphorylation take place *in vivo*?

Phosphorylation of two threonine residues of RBPMS regulated its splicing activity as indicated by mutations that mimicked phosphorylation (Chapter 7). However, a major remaining question is whether these post-translation modifications indeed happen in the cell. These modifications were reported by proteomic studies (Hornbeck et al., 2015)

but were not validated in this study. Nevertheless, treatment of protein lysates from PAC1 cells with phosphatases caused a protein size shift in RBPMS when visualized in a polyacrylamide gel, indicating that RBPMS is potentially phosphorylated in these cells (experiment carried out by C. Gooding in the laboratory). Yet this does not provide direct evidence for the specific modification of the Thr 113 and/or 118. Moreover, the molecular mechanism behind inhibition of RBPMS splicing by these modifications is not clear. Due to the proximity of these residues to the RRM, it is likely that they disrupt RNA-binding, but that should affect both repression and activation, and not either one. Altered localization or dimerization also cannot explain inhibition of a specific activity while maintaining the other. Future approaches to be applied to better understand the role of phosphorylation in the regulation of RBPMS as well as the conditions in which it occurs are discussed below.

8.1.5 Does RBPMS liquid phase separate?

In view of the fact that the recombinant RBPMS seems to assemble into something much larger than just a dimer (Chapter 5-6), one could wonder if this relates to its activity in the cell. Indeed, RBFOX proteins have been shown to assemble into large complexes (Damianov et al., 2016; Ying et al., 2017) and the potential similarities with RBPMS were raised in Chapter 7 (see Discussion section). Despite RBP aggregation being associated with pathological conditions in the past, a very topical subject in the field of RBPs is the physiological roles of these "aggregates" (Polymenidou, 2018). In fact, actual aggregates are insoluble aberrant assemblies and more fibrous-like structures whereas these recently characterized "aggregates" consist of dynamic and reversible assemblies. Therefore these physiological "aggregates" have been more correctly termed as higher-order assemblies (Polymenidou, 2018). The dynamic nature of these assemblies allows them to exhibit a phase transition behavior (Polymenidou, 2018). This liquid-liquid phase separation promotes compartmentalization of functions in membrane-less organelles-like structures (RNP granules) (Lin et al., 2015). In the cytoplasm, self-assembly is involved in the formation of cytoplasmic stress granules and transport granules whereas in the nucleus it is believed that the self-assembly concentrates factors in the nucleolus and the Cajal body (Lin et al., 2015; Polymenidou, 2018). The self-assembly property can be conferred by intrinsically disordered regions of RBPs and recently shown to regulate splicing by RBFOX (Ying et al., 2017). The splicing capability of the tyrosine repetitive sequence in the disordered region of RBFOX was associated with its ability to self-assemble and localize within the nucleus. Thus, the intriguing formation of large complexes by RBPMS recombinant protein

and the presence of aggregation features in its C-terminal disordered domain hint that higher-order assemblies might be a feature of RBPMS splicing regulation. Indeed, self-assembly could be relevant to the observation that RBPMS is associated with stress granules in the cytoplasm (Farazi et al., 2014).

8.2 Conclusion

This study provided clear evidence that RBPMS acts as a master splicing regulator of VSMCs. It also vindicated Jangi and Sharp's suggested approach for the identification of tissue-specific master splicing regulators. By focusing on RBP genes associated with SM super-enhancers, RBPMS was identified as a key regulator of the differentiated SMC AS program. RBPMS was predominantly nuclear and found expressed in the rat pulmonary artery PAC1 cells in two major isoforms: RBPMS-A and RBPMS-B, which vary in their extreme C-termini. Initial experiments with SMC minigene reporters (*Tpm1* and *Actn1*) supported the role of RBPMS in AS of differentiated PAC1 cells.

Based on these results, mRNA-Seq was carried out for RBPMS knockdown and RBPMS-A overexpression in differentiated and dedifferentiated PAC1 cells respectively. RBPMS promoted numerous differentiated splicing patterns, solely regulating 20% of the AS changes during PAC1 phenotypic switching. Additionally, RBPMS-A overexpression was sufficient to promote fully differentiated SMC tissue-like AS patterns, as exemplified by *Fermt2*, *Cald1* and *Tsc2*. RBPMS targeted a network of proteins involved in the actin cytoskeleton, focal adhesions and the secretory pathway: all critical machineries in the interconversion between the contractile differentiated and the motile-synthetically active dedifferentiated SMC states. Super-enhancer association of RBPMS targets further reinforced the importance of these genes to the SMC biology.

Bioinformatic tools and biochemical assays confirmed RBPMS direct regulation of splicing. Similarly to other AS regulators, RBPMS splicing activity was position-dependent; sites upstream and within exon associated with repressed exons whereas downstream sites with activated exons as indicated by RBPMS splicing maps. Moreover, *in vivo* transfections with wild-type and CAC mutant minigene reporters as well as RBPMS RNA-binding mutant and lastly *in vitro* binding assays (EMSA and UV-crosslinking) consistently supported RBPMS direct regulation of SMC splicing via CAC motifs. Nevertheless, RBPMS could also lead to secondary effects on splicing and transcription by controlling splicing and activity of other regulators of AS (*Mbnl1, Mbnl2*), mRNA stability (*Lsm14b*) as well as a key SMC transcription factor (*Myocd*). Notably, RBPMS promoted production of a MYOCD isoform that more potently activates the contractile phenotype, acting antagonistically with QKI which is more highly expressed in proliferative cells.

In addition to that, a few features underlining RBPMS splicing activity were provided in this work by applying structure-function studies. RBPMS dimerization was essential for its activity, although RBPMS RRM was not sufficient to promote splicing changes. Moreover, RBPMS N-terminus was dispensable for its activity whereas the

C-terminus was required for its function. Surprisingly, a core section of the RBPMS-B C-terminus antagonized the repressor activity, potentially explaining the lower activity of RBPMS-B compared to RBPMS-A, particularly for splicing repression. Moreover, two Thr residues (position 113 and 118) of RBPMS adjacent to its RRM domain could be phosphorylated. Phosphomimetic mutants differentially modulated RBPMS activity, inhibiting splicing repression by RBPMS-A and splicing activation by RBPMS-B.

In summary, this study provides the strongest evidence to date for a molecular function of RBPMS as a splicing regulator, matching many of the expected criteria of a master regulator of AS in differentiated VSMCs: a high dynamic range of expression during SMC dedifferentiation; regulation of a coherent set of targets important for SMC function; and regulation of other post-transcriptional and transcriptional regulators. Finally, RBPMS splicing activity was further characterized in this study uncovering some specific mechanisms of its AS regulation and differential activity of RBPMS isoforms. The main findings of this work are listed and summarized in the following model (Figure 8.3).

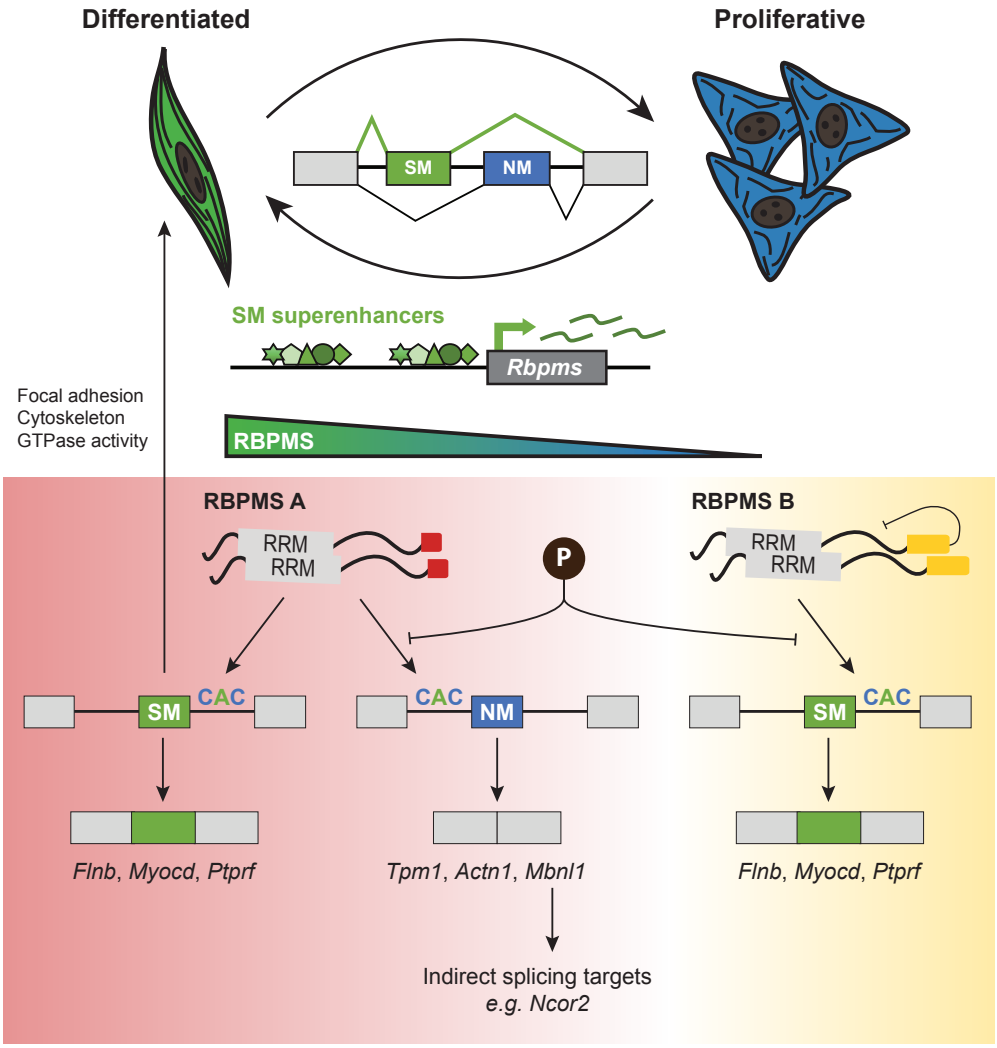


Fig. 8.3 Summary of the main findings of this study that corroborate RBPMS as a smooth muscle master splicing regulator.

Highlights

- **Super-enhancer association to identify master splicing regulators**
- **RBPMS is an alternative splicing regulator**
- **Differential activity of RBPMS isoforms**
- **RBPMS targets mRNAs important for SMC function (Figure 8.4)**
- **RBPMS splicing activity requires additional sequences to its RRM**
- **RBPMS splicing cascade involves other regulators (MBNL and MYOCD)**
- **RBPMS phosphorylation modulates its splicing activity**

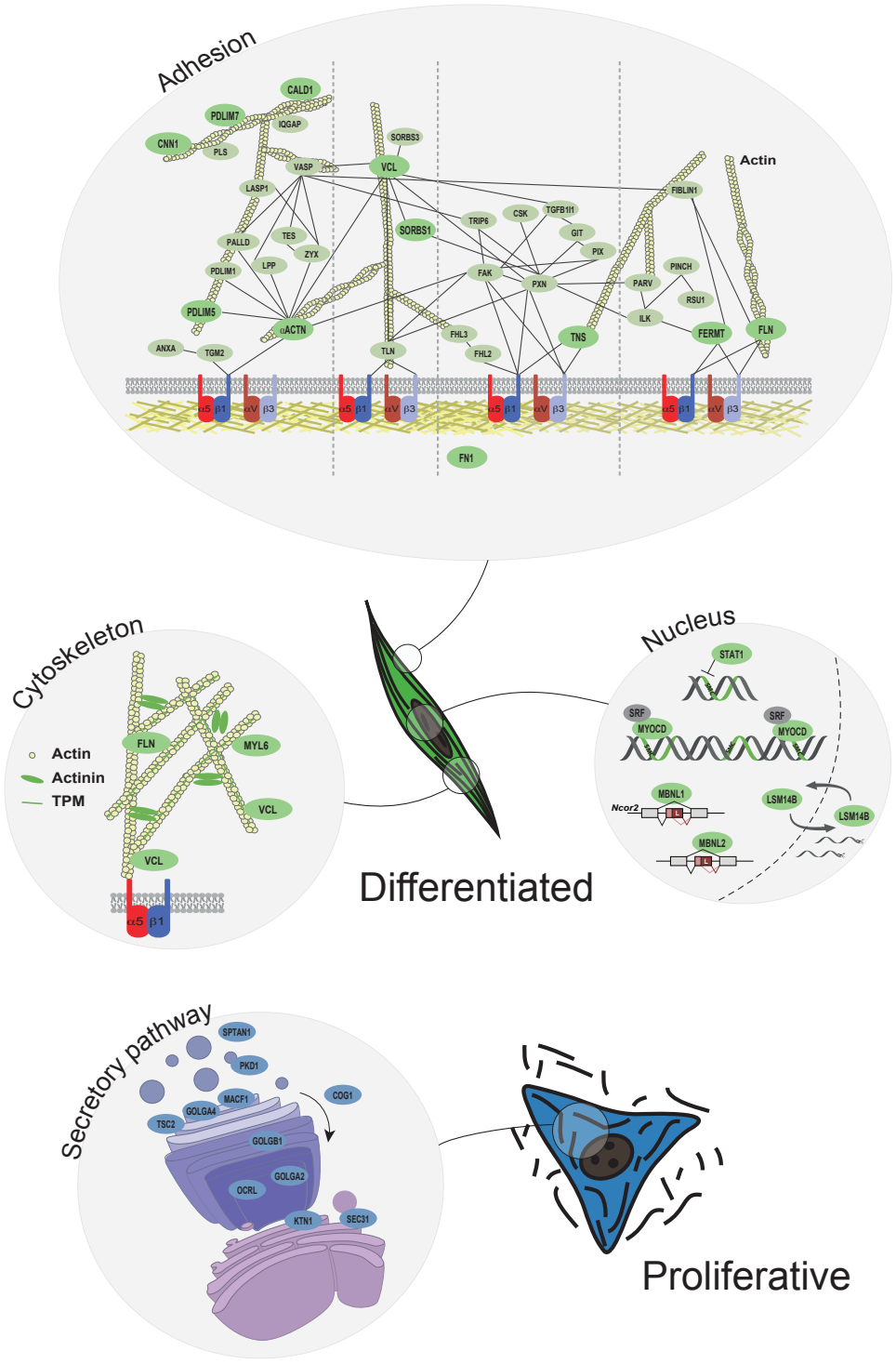


Fig. 8.4 SMC functions targeted by RBPMS-mediated alternative splicing. RBPMS-mediated AS affected SMC genes involved in roles in the nucleus, cytoskeleton, cell-adhesion and secretion as highlighted in dark green and blue.

8.3 Future directions

Identification of RBPMS as a splicing regulator of VSMCs led to further questions reported throughout this work. Although some of these have been addressed in here, some questions remain to be answered in future studies (Figure 8.5).

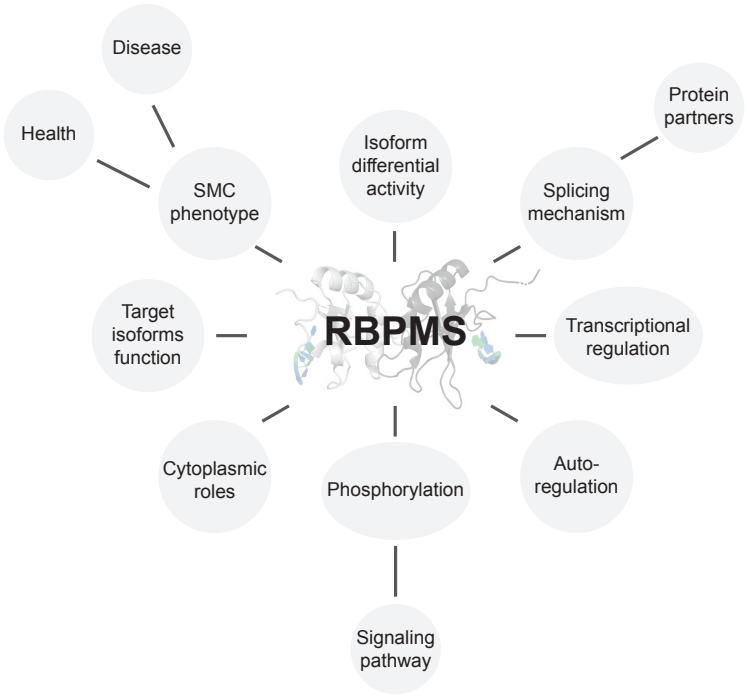


Fig. 8.5 Schematic of the future directions for better understanding of RBPMS.

RBPMS regulated splicing in controlling SMC phenotype

Despite the regulation of hundreds of ASEs, of which some have well-characterized functions, manipulation of RBPMS levels, either by depletion or overexpression, did not lead to any apparent phenotype in the PAC1 cells. Due to the involvement of the genes regulated by RBPMS in cytoskeleton and focal adhesion functions, a few phenotypic screenings focusing on those functions were carried out but no striking alterations in the arrangement of actin fibers or size of the focal adhesion were observed. Moreover, migratory capacities were not altered upon RBPMS knockdown or overexpression. Nonetheless, these experiments need to be repeated and further analyzed into more details to unravel any phenotype. Additionally, the cultured PAC1 system might not consist of the most appropriate model and despite the fact that PAC1 cells retain some of the differentiation features of SMC, other models representative of the more

differentiated state could be explored such as primary cells or even generation of a transgenic animal model to get insights into function of RBPMS in VSMCs.

SMCs also show a great deal of diversity (Fisher, 2010), even within single blood vessels (Cheung et al., 2012). Therefore, an interesting scenario could arise from the association of RBPMS with more differentiated cells, a feature already highlighted in a single cell RNA-Seq in which *Rbpms* was identified as part of a transcriptome signature of contractile mouse aorta SMCs cells (Dobnikar et al., 2018). Moreover, diversity is also found in the splicing of tonic and phasic SMCs (Shukla and Fisher, 2008). Thus, even though RBPMS was identified regulating the VSMC splicing program, other RBPs might act as master regulators of some of these specialized SMC types.

Finally, the role RBPMS in SMC phenotypic switching could also be investigated in disease conditions, when the SMC interconversion is dysregulated (Bennett et al., 2016; Frismantiene et al., 2018). Indeed, RBPMS levels have been shown to decrease in the carotid ligated mouse vessels (unpublished data from Dr. Aishwarya Jacob), indicating a potential role of RBPMS in the maintenance of healthy vascular tissues. Additionally, the carotid injury model could be used in the future to look at AS changes during phenotypic modulation *in vivo*. Another interesting possibility that could associate RBPMS with cardiovascular diseases is that disease associated SNPs could create or remove CAC motifs that then dysregulate RBPMS-regulated ASEs.

RBPMS isoform differential activity

Experiments carried out in here revealed that RBPMS activity was regulated via generation of different isoforms via AS of its own transcript. However, mRNAseq was only carried out for PAC1 cells overexpressing RBPMS-A. Taken the differential splicing activity of RBPMS isoforms, it would be interesting to also investigate the global targets of RBPMS-B. This would provide a better dataset of common and distinct events regulated by RBPMS isoforms. Moreover, it could reveal further biological implications of different levels of expression of the isoforms in SMCs. In that way, approaches for manipulation of individual isoforms by either CRISPR-Cas9 or ASO could be carried out in SMCs.

In fact, regulation of the alternative 3′ end exon which determines the two major RBPMS isoforms should also be investigated. Inclusion of RBPMS exon 7 as opposed to exon 8 leads to expression of RBPMS-B, which is the least active isoform. Thus, insights could be gained by the study of the aspects of these exons affecting this critical splicing decision in RBPMS. In addition to that, RBPMS exon 6 skipping produced several non-functional isoforms which could consist of a negative feedback loop in

the regulation of RBPMS levels. Interestingly, clusters of RBPMS optimal binding sites (CAC motifs) could be identified downstream of exon 6, indicating a potential auto-regulatory mechanism.

Mechanism of splicing regulation by RBPMS

In order to address the different properties of RBPMS-A and RBPMS-B determining their repressive activity, additional biophysical studies have been carried out in the laboratory, especially focusing on their capability to form large assemblies (data not shown). Proteomic studies by co-immunoprecipitations and mass spectrometry should be aimed to uncover any potential protein-protein interaction with co-regulatory proteins and/or target core splicing factors that could explain RBPMS splicing function and differential activity. Interestingly, STRING generated RBPMS interactome shows interactions with other RBPs involved in mRNA splicing, e.g. RBFOX2, QKI and ESRP1 (Figure 8.6). Epithelial Splicing Regulatory Protein 1 (ESRP1) was identified to interact with RBPMS2 via immunoprecipitation in chicken DF1 fibroblast cell line (Sagnol et al., 2016). While interactions with other RBPs were identified in large scale studies and have not been analyzed in detail.

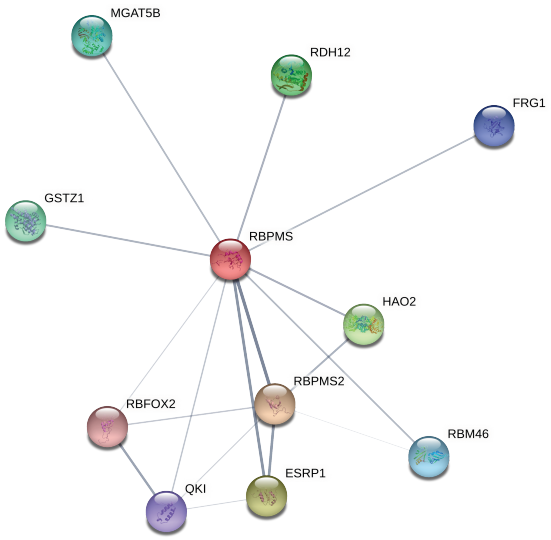


Fig. 8.6 Human RBPMS interactome RBPMS protein partners interactome in humans generated using STRING (v 11.0) and using experiments as the only source of interactions. Network edges mean confidence where line thickness indicate strength of data support

Another methodology to be applied in the characterization of RBPMS splicing activity is MS2 artificial tethering. In this study (Chapter 7), initial experiments were

carried out establishing MS2 assays in repressed and activated exons (*Tpm1* exon 3 and *Myocd* exon 2a). Even though only RBPMS full length and RBPMS-A binding mutant were tested in the assays shown here, a library of several RBPMS truncated mutants fused to an MS2 domain has been generated to be used in the identification of RBPMS "effector" regions. These mutants could unravel regions required for RBPMS splicing in addition to RNA-binding and dimerization conferred by the RRM. Thus, MS2-tethering will test whether recruited effector domains are sufficient to mediate regulation. Lastly, MS2-tethering studies will be applied to investigate the role of the isoform specific CTDs.

Interestingly, nuclear extracts from HEK293 overexpressing RBPMS-A were sufficient to repress *Tpm1* exon 3 in an *in vitro* splicing assay (Appendix - Figure A.7 and A.8, data from Fredderick Richards, Part III student). Therefore, by further applying this method, more information about RBPMS splicing mechanism could be gained in the future. Additionally, taken that both isoforms, A and B, were able to bind to *Tpm1* RNA, *in vitro* splicing assays could also highlight the different steps of splicing in which RBPMS-B fails to promote *Tpm1* repression. Eventually, regulatory complexes could be assembled on *Tpm1* transcripts with exon 3 and flanking regulatory elements in nuclear extracts with and without RBPMS. The high resolution structure of these macromolecular complexes could then be assessed by CryoEM to provide rich insights into RBPMS mechanism of action.

RBPMS splicing cascade: indirect targets

RBPMS regulated other post-transcriptional factors involved in mRNA splicing, like *Mbnl1* and *Mbnl2*, and in other processes such as mRNA stability by regulation of *Lsm14b*. Consequently, RBPMS was associated with indirect splicing changes by the switch of MBNL isoforms, e.g. *Ncor2* A5SS. Therefore, in order to distinguish direct and secondary targets of RBPMS in a more high-throughput manner, iCLIP experiments could be preformed in the SMCs. RBPMS iCLIP experiments would allow elucidation of the RNA-RBPMS interactions taking place in these cells and could also be used to address if there is any difference in the global binding between RBPMS isoforms. These datasets could further vindicate RBPMS regulation of splicing by direct binding to pre-mRNAs *in vivo* and shed light onto indirect splicing events.

RBPMS also controlled splicing of a SM transcription factor, *Myocd*. Although no changes in SMC markers were observed upon manipulation of RBPMS levels, regulation of such a transcription-factor is expected to modulate expression of differentiation markers, leading to greater contractility. The lack of MYOCD activation of SMC genes

could be explained by the limitations imposed by the transient nature of RBPMS knockdown and overexpression in PAC1 cells. So in the future, RBPMS knockout by CRISPR-Cas9 and sustained RBPMS overexpression could provide better insights into the regulation of the differentiation state upon RBPMS splicing control of *Myocd* transcripts among other targets.

Modulation of RBPMS activity by post-translational modification

Investigation of the regulation of RBPMS by PTMs revealed phosphorylation as a major regulator of its activity. However, neither the signaling pathway nor the kinase/phosphatase involved in RBPMS phosphorylation were found out in this study. Thereby they will be the aim of future experiments. These studies could uncover potential environmental cues that are involved in the activation or inhibition of RBPMS-mediated splicing. Moreover, the full understanding of the interplay between the kinase, RBPMS and its RNA targets may provide novel insights into the dynamic reshaping of SMC transcriptome by RBPMS.

To understand how phosphorylation regulates RBPMS activity, proteomic studies could also include T113/118 phospho-site mutants. Protein interactions disrupted or acquired by these phosphorylated residues could potentially be identified. Finally, phosphorylation could allosterically control RBPMS activity due to the proximity of the Thr to RBPMS RRM domain. Thereby, structural studies could also be explored to resolve the mechanism behind RBPMS phosphorylation.

Cytoplasmic roles of RBPMS

In this work, the focus was given to the nuclear role of RBPMS in post-transcriptional processes, specifically in alternative splicing. However, most attention has been paid to RBPMS cytoplasmic roles in the studies investigating RBPMS. In fact RBPMS has been implicated in several other mRNA processes by these studies (Farazi et al., 2014; Furukawa et al., 2015; Hörnberg et al., 2013; Rambout et al., 2016; Sagnol et al., 2016). Therefore, there is a lot of scope for RBPMS functions in mRNA stability, localization and translation, whose importance in SMCs is yet to be determined.

Transcriptional control of RBPMS

In this study, RBPMS was found acting as a master regulator of VSMCs, reshaping the transcriptome of these cells. Yet *Rbpms* is still under transcriptional control. In addition to *Rbpms* association with SMC super-enhancers, only one single study attempted addressing RBPMS transcriptional control (Ye et al., 2018). The knowledge

of the transcription factors driving *Rbpms* expression could contribute to the better understanding of the strong downregulation of *Rbpms* levels during SMC dedifferentiation. RBPMS levels could also be titrated in some conditions by non-coding RNAs (ncRNAs). For instance, lncRNAs (long noncoding RNAs) containing repetitive CAC sequences could sequester RBPMS and control its splicing activity, in a similar way to PTBP1 and the lncRNA PNCTR (Yap et al., 2018). Thus, investigation of RBPMS expression control and titration by ncRNAs would provide insights into its distinct expression pattern and activity in different tissues and also elucidate its downregulation in some disorders involving SMC phenotypic plasticity, like cardiovascular diseases.

SMC splicing code

Tissue-specific AS programs are likely to arise from the combinatorial effects of the different *trans*-factors. In this study, we addressed the interplay between RBPMS and QKI, a known regulator of the proliferative state, in the regulation of some common AS targets (*Myocd* and *Flnb*). These factors were as expected antagonistic, but QKI seemed to be the dominant regulator. Additionally, PTBP1 and MBNL are also known to regulate *Tpm1* and initial experiments by Fredderick Richards have also revealed that these factors are necessary for full repression of *Tpm1* exon 3 by RBPMS (data not shown). Therefore, future studies will also aim to characterize the different input signals from cooperative and antagonistic interactions between RBPMS, a master regulator of the differentiated VSMCs, and regulators of the proliferative state, e.g. QKI and PTBP1, in RBPMS splicing decisions.

In conclusion, future studies will aim to understand the mechanism of splicing regulation by RBPMS, the role of the RBPMS regulated splicing program in controlling different aspects of SMC phenotype and finally, the potential role of subversion of this program in cardiovascular diseases.

References

Aguero, T., Zhou, Y., Kloc, M., Chang, P., Houliston, E., and King, M. L. (2016). Hermes (Rbpms) is a Critical Component of RNP Complexes that Sequester Germline RNAs during Oogenesis. *Journal of developmental biology*, 4(1):2.

Alt, F. W., Bothwell, A. L., Knapp, M., Siden, E., Mather, E., Koshland, M., and Baltimore, D. (1980). Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. *Cell*, 20(2):293–301.

Amara, S. G., Jonas, V., Rosenfeld, M. G., Ong, E. S., and Evans, R. M. (1982). Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. *Nature*, 298(5871):240–244.

Ascano, M., Hafner, M., Cekan, P., Gerstberger, S., and Tuschl, T. (2012). Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdisciplinary Reviews: RNA*, 3(2):159–177.

Ast, G. (2004). How did alternative splicing evolve? *Nature Reviews Genetics*, 5(10):773–782.

Auweter, S. D., Fasan, R., Reymond, L., Underwood, J. G., Black, D. L., Pitsch, S., and Allain, F. H.-T. (2006). Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *The EMBO Journal*, 25(1):163–173.

Baek, D. and Green, P. (2005). Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. *Proceedings of the National Academy of Sciences*, 102(36):12813–12818.

Báez-Vega, P. M., Vargas, I. M. E., Valiyeva, F., Encarnación-Rosado, J., Roman, A., Flores, J., Marcos-Martínez, M. J., and Vivas-Mejía, P. E. (2016). Targeting miR-21-3p inhibits proliferation and invasion of ovarian cancer cells. *Oncotarget*, 7(24):36321–36337.

Bar-Sagi, D. and Hall, A. (2000). Ras and Rho GTPases: a family reunion. *Cell*, 103(2):227–38.

Baralle, F. E. and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nature Reviews Molecular Cell Biology*, 18(7):437–451.

Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., Kim, T., Misquitta-Ali, C. M., Wilson, M. D., Kim, P. M., Odom, D. T., Frey, B. J., and Blencowe, B. J. (2012).

The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. *Science*, 338(6114):1587–1593.

Batra, R., Charizanis, K., Manchanda, M., Mohan, A., Li, M., Finn, D. J., Goodwin, M., Zhang, C., Sobczak, K., Thornton, C. A., and Swanson, M. S. (2014). Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Molecular cell*, 56(2):311–322.

Bennett, M. R., Sinha, S., and Owens, G. K. (2016). Vascular Smooth Muscle Cells in Atherosclerosis. *Circulation Research*, 118(4):692–702.

Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics*, 15(3):163–175.

Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74(8):3171–5.

Bertram, K., Agafonov, D. E., Dybkov, O., Haselbach, D., Leelaram, M. N., Will, C. L., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2017). Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell*, 170(4):701–713.e11.

Beuck, C., Qu, S., Fagg, W. S., Ares, M., and Williamson, J. R. (2012). Structural analysis of the quaking homodimerization interface. *Journal of molecular biology*, 423(5):766–81.

Binder, J. X., Pletscher-Frankild, S., Tsafou, K., Stolte, C., O'Donoghue, S. I., Schneider, R., and Jensen, L. J. (2014). COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database*, 2014(0):bau012–bau012.

Black, D. L. (2003). Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry*, 72(1):291–336.

Blencowe, B. J. (2006). Alternative Splicing: New Insights from Global Analyses. *Cell*, 126(1):37–47.

Boettger, T., Beetz, N., Kostin, S., Schneider, J., Krüger, M., Hein, L., and Braun, T. (2009). Acquisition of the contractile phenotype by murine arterial smooth muscle cells depends on the Mir143/145 gene cluster. *Journal of Clinical Investigation*, 119(9):2634–2647.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120.

Boutz, P. L., Chawla, G., Stoilov, P., and Black, D. L. (2007a). MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. *Genes & Development*, 21(1):71–84.

Boutz, P. L., Stoilov, P., Li, Q., Lin, C.-H., Chawla, G., Ostrow, K., Shiue, L., Ares, M., and Black, D. L. (2007b). A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes & Development*, 21(13):1636–1652.

Brandmann, T., Fakim, H., Padamsi, Z., Youn, J.-Y., Gingras, A.-C., Fabian, M. R., and Jinek, M. (2018). Molecular architecture of LSM14 interactions involved in the assembly of mRNA silencing complexes. *The EMBO journal*, 37(7):e97869.

Buljan, M., Chalancon, G., Eustermann, S., Wagner, G. P., Fuxreiter, M., Bateman, A., and Babu, M. M. (2012). Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Molecular cell*, 46(6):871–83.

Calabretta, S. and Richard, S. (2015). Emerging Roles of Disordered Sequences in RNA-Binding Proteins. *Trends in Biochemical Sciences*, 40(11):662–672.

Caputi, M. and Zahler, A. M. (2001). Determination of the RNA Binding Specificity of the Heterogeneous Nuclear Ribonucleoprotein (hnRNP) H/H'/F/2H9 Family. *Journal of Biological Chemistry*, 276(47):43850–43859.

Castello, A., Fischer, B., Frese, C. K., Horos, R., Alleaume, A.-M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M. W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular cell*, 63(4):696–710.

Cheung, C., Bernardo, A. S., Pedersen, R. A., and Sinha, S. (2014). Directed differentiation of embryonic origin–specific vascular smooth muscle subtypes from human pluripotent stem cells. *Nature Protocols*, 9(4):929–938.

Cheung, C., Bernardo, A. S., Trotter, M. W. B., Pedersen, R. A., and Sinha, S. (2012). Generation of human vascular smooth muscle subtypes provides insight into embryological origin–dependent disease susceptibility. *Nature Biotechnology*, 30(2):165–173.

Chintalapudi, S. R., Djenderedjian, L., Stiemke, A. B., Steinle, J. J., Jablonski, M. M., and Morales-Tirado, V. M. (2016). Isolation and Molecular Profiling of Primary Mouse Retinal Ganglion Cells: Comparison of Phenotypes from Healthy and Glaucomatous Retinas. *Frontiers in Aging Neuroscience*, 8:93.

Chow, L. T. and Broker, T. R. (1978). The spliced structures of adenovirus 2 fiber message and the other late mRNAs. *Cell*, 15(2):497–510.

Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, 12(1):1–8.

Climente-González, H., Porta-Pardo, E., Godzik, A., and Eyras, E. (2017). The Functional Impact of Alternative Splicing in Cancer. *Cell Reports*, 20(9):2215–2226.

Conchillo-Solé, O., de Groot, N. S., Avilés, F. X., Vendrell, J., Daura, X., and Ventura, S. (2007). AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics*, 8(1):65.

Conlon, E. G. and Manley, J. L. (2017). RNA-binding proteins in neurodegeneration: mechanisms in aggregate. *Genes & development*, 31(15):1509–1528.

Cordes, K. R., Sheehy, N. T., White, M. P., Berry, E. C., Morton, S. U., Muth, A. N., Lee, T.-H., Miano, J. M., Ivey, K. N., and Srivastava, D. (2009). miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. *Nature*, 460(7256):705–10.

Creemers, E. E., Sutherland, L. B., Oh, J., Barbosa, A. C., and Olson, E. N. (2006). Coactivation of MEF2 by the SAP Domain Proteins Myocardin and MASTR. *Molecular Cell*, 23(1):83–96.

Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, 227(5258):561–563.

Crick, F. H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology*, 12:138–63.

Damianov, A. and Black, D. L. (2010). Autoregulation of Fox protein expression to produce dominant negative splicing factors. *RNA*, 16(2):405–16.

Damianov, A., Ying, Y., Lin, C.-H., Lee, J.-A., Tran, D., Vashisht, A., Bahrami-Samani, E., Xing, Y., Martin, K., Wohlschlegel, J., and Black, D. (2016). Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. *Cell*, 165(3):606–619.

Daubner, G. M., Cléry, A., and Allain, F. H.-T. (2013). RRM–RNA recognition: NMR or crystallography... and new findings. *Current Opinion in Structural Biology*, 23(1):100–108.

Davis-Dusenbery, B. N., Wu, C., Hata, A., and Sessa, W. C. (2011). Micromanaging Vascular Smooth Muscle Cell Differentiation and Phenotypic Modulation. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 31(11):2370–2377.

de Bruin, R. G., Rabelink, T. J., van Zonneveld, A. J., and van der Veer, E. P. (2017). Emerging roles for RNA-binding proteins as effectors and regulators of cardiovascular disease. *European heart journal*, 38(18):1380–1388.

Deribe, Y. L., Pawson, T., and Dikic, I. (2010). Post-translational modifications in signal integration. *Nature Structural & Molecular Biology*, 17(6):666–672.

DiDonato, J., Mercurio, F., Rosette, C., Wu-Li, J., Suyang, H., Ghosh, S., and Karin, M. (1996). Mapping of the inducible IkappaB phosphorylation sites that signal its ubiquitination and degradation. *Molecular and cellular biology*, 16(4):1295–304.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21.

Dobnikar, L., Taylor, A. L., Chappell, J., Oldach, P., Harman, J. L., Oerton, E., Dzierzak, E., Bennett, M. R., Spivakov, M., and Jørgensen, H. F. (2018). Disease-relevant transcriptional signatures identified in individual smooth muscle cells from healthy mouse vessels. *Nature Communications*, 9(1):4567.

Dvinge, H. and Bradley, R. K. (2015). Widespread intron retention diversifies most cancer transcriptomes. *Genome Medicine*, 7(1):45.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., and Hood, L. (1980). Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. *Cell*, 20(2):313–9.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, 10(1):48.

El Marabti, E. and Younis, I. (2018). The Cancer Spliceome: Reprograming of Alternative Splicing in Cancer. *Frontiers in Molecular Biosciences*, 5:80.

Eperon, L. P., Graham, I. R., Griffiths, A. D., and Eperon, I. C. (1988). Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell*, 54(3):393–401.

Fagg, W. S., Liu, N., Fair, J. H., Shiue, L., Katzman, S., Donohue, J. P., and Ares, M. (2017). Autogenous cross-regulation of Quaking mRNA processing and translation balances Quaking functions in splicing and translation. *Genes & development*, 31(18):1894–1909.

Farazi, T. A., Leonhardt, C. S., Mukherjee, N., Mihailovic, A., Li, S., Max, K. E. A., Meyer, C., Yamaji, M., Cekan, P., Jacobs, N. C., Gerstberger, S., Bognanni, C., Larsson, E., Ohler, U., and Tuschl, T. (2014). Identification of the RNA recognition element of the RBPMS family of RNA-binding proteins and their transcriptome-wide mRNA targets. *RNA*, 20(7):1090–1102.

Fernandez-Costa, J. M., Llamusi, M. B., Garcia-Lopez, A., and Artero, R. (2011). Alternative splicing regulation by Muscleblind proteins: from development to disease. *Biological Reviews*, 86(4):947–958.

Fica, S. M. and Nagai, K. (2017). Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nature Structural & Molecular Biology*, 24(10):791–799.

Fisette, J.-F., Toutant, J., Dugre-Brisson, S., Desgroseillers, L., and Chabot, B. (2010). hnRNP A1 and hnRNP H can collaborate to modulate 5' splice site selection. *RNA*, 16(1):228–238.

Fisher, S. A. (2010). Vascular smooth muscle phenotypic diversity and function. *Physiological Genomics*, 42A(3):169–187.

Fogel, B. L., Wexler, E., Wahnich, A., Friedrich, T., Vijayendran, C., Gao, F., Parikshak, N., Konopka, G., and Geschwind, D. H. (2012). RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Human Molecular Genetics*, 21(19):4171–4186.

Förch, P., Puig, O., Martínez, C., Séraphin, B., and Valcárcel, J. (2002). The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. *The EMBO journal*, 21(24):6882–92.

Frismantiene, A., Philippova, M., Erne, P., and Resink, T. J. (2018). Smooth muscle cell-driven vascular diseases and molecular mechanisms of VSMC plasticity. *Cellular Signalling*, 52:48–64.

Fu, J., Cheng, L., Wang, Y., Yuan, P., Xu, X., Ding, L., Zhang, H., Jiang, K., Song, H., Chen, Z., and Ye, Q. (2015). The RNA-binding protein RBPMS1 represses AP-1 signaling and regulates breast cancer cell proliferation and migration. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1853(1):1–13.

Fu, X.-D. and Ares, M. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nature Reviews Genetics*, 15(10):689–701.

Furukawa, M. T., Sakamoto, H., and Inoue, K. (2015). Interaction and colocalization of HERMES/RBPMS with NonO, PSF, and G3BP1 in neuronal cytoplasmic RNP granules in mouse retinal line cells. *Genes to Cells*, 20(4):257–266.

Gardina, P. J., Clark, T. A., Shimada, B., Staples, M. K., Yang, Q., Veitch, J., Schweitzer, A., Awad, T., Sugnet, C., Dee, S., Davies, C., Williams, A., and Turpaz, Y. (2006). Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics*, 7(1):325.

Garg, K. and Green, P. (2007). Differing patterns of selection in alternative and constitutive splice sites. *Genome research*, 17(7):1015–22.

Gerber, W. V., Vokes, S. a., Zearfoss, N. R., and Krieg, P. a. (2002). A role for the RNA-binding protein, hermes, in the regulation of heart development. *Developmental biology*, 247(1):116–26.

Gerber, W. V., Yatskievych, T. A., Antin, P. B., Correia, K. M., Conlon, R. A., and Krieg, P. A. (1999). The RNA-binding protein gene, hermes, is expressed at high levels in the developing heart. *Mechanisms of Development*, 80(1):77–86.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nature Reviews Genetics*, 15(12):829–845.

Gohr, A. and Irimia, M. (2018). Matt: Unix tools for alternative splicing analysis. *Bioinformatics*.

Gooding, C., Edge, C., Lorenz, M., Coelho, M. B., Winters, M., Kaminski, C. F., Cherny, D., Eperon, I. C., and Smith, C. W. J. (2013). MBNL1 and PTB cooperate to repress splicing of Tpm1 exon 3. *Nucleic Acids Research*, 41(9):4765–4782.

Gooding, C., Roberts, G. C., and Smith, C. W. (1998). Role of an inhibitory pyrimidine element and polypyrimidine tract binding protein in repression of a regulated alpha-tropomyosin exon. *RNA*, 4(1):85–100.

Gooding, C. and Smith, C. W. J. (2008). Tropomyosin exons as models for alternative splicing. *Advances in experimental medicine and biology*, 644:27–42.

Groen, E. J. N., Talbot, K., and Gillingwater, T. H. (2018). Advances in therapy for spinal muscular atrophy: promises and challenges. *Nature Reviews Neurology*, 14(4):214–224.

Gromak, N., Matlin, A. J., Cooper, T. A., and Smith, C. W. J. (2003). Antagonistic regulation of alpha-actinin alternative splicing by CELF proteins and polypyrimidine tract binding protein. *RNA*, 9(4):443–456.

Gromak, N. and Smith, C. W. J. (2002). A splicing silencer that regulates smooth muscle specific alternative splicing is active in multiple cell types. *Nucleic acids research*, 30(16):3548–57.

Gueroussov, S., Weatheritt, R. J., O'Hanlon, D., Lin, Z.-Y., Narula, A., Gingras, A.-C., and Blencowe, B. J. (2017). Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing. *Cell*, 170(2):324–339.e23.

Hamid, F. M. and Makeyev, E. V. (2017). A mechanism underlying position-specific regulation of alternative splicing. *Nucleic acids research*, 45(21):12455–12468.

Hentze, M. W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology*, 19(5):327–341.

Herzel, L., Ottoz, D. S. M., Alpert, T., and Neugebauer, K. M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nature reviews. Molecular cell biology*, 18(10):637–650.

Heyd, F. and Lynch, K. W. (2011). DEGRADE, MOVE, REGROUP: signaling control of splicing proteins. *Trends in Biochemical Sciences*, 36(8):397–404.

Hnisz, D., Abraham, B., Lee, T., Lau, A., Saint-André, V., Sigova, A., Hoke, H., and Young, R. (2013). Super-Enhancers in the Control of Cell Identity and Disease. *Cell*, 155(4):934–947.

Hnisz, D., Schuijers, J., Lin, C. Y., Weintraub, A. S., Abraham, B. J., Lee, T. I., Bradner, J. E., and Young, R. A. (2015). Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Molecular cell*, 58(2):362–70.

Hofweber, M. and Dormann, D. (2018). Friend or foe - post-translational modifications as regulators of phase separation and RNP granule dynamics. *The Journal of biological chemistry*, page jbc.TM118.001189.

Hollander, D., Naftelberg, S., Lev-Maor, G., Kornblihtt, A. R., and Ast, G. (2016). How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing? *Trends in genetics : TIG*, 32(10):596–606.

Honkaniemi, J., Zhang, J. S., Yang, T., Zhang, C., Tisi, M. A., and Longo, F. M. (1998). LAR tyrosine phosphatase receptor: proximal membrane alternative splicing is coordinated with regional expression and intraneuronal localization. *Brain research. Molecular brain research*, 60(1):1–12.

Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic acids research*, 43(Database issue):D512–20.

Hörnberg, H., Cioni, J.-M., Harris, W. A., and Holt, C. E. (2016). Hermes Regulates Axon Sorting in the Optic Tract by Post-Trancriptional Regulation of Neuropilin 1. *Journal of neuroscience*, 36(50):12697–12706.

Hörnberg, H., Wollerton-van Horck, F., Maurus, D., Zwart, M., Svoboda, H., Harris, W. A., and Holt, C. E. (2013). RNA-binding protein Hermes/RBPMS inversely affects synapse density and axon arbor formation in retinal ganglion cells in vivo. *Journal of neuroscience*, 33(25):10384–95.

Horton, E. R., Byron, A., Askari, J. A., Ng, D. H. J., Millon-Frémillon, A., Robertson, J., Koper, E. J., Paul, N. R., Warwood, S., Knight, D., Humphries, J. D., and Humphries, M. J. (2015). Definition of a consensus integrin adhesome and its dynamics during adhesion complex assembly and disassembly. *Nature Cell Biology*, 17(12):1577–1587.

Hou, N., Guo, Z., Zhao, G., Jia, G., Luo, B., Shen, X., and Bai, Y. (2018). Inhibition of microRNA-21-3p suppresses proliferation as well as invasion and induces apoptosis by targeting RNA-binding protein with multiple splicing through Smad4/extra cellular signal-regulated protein kinase signalling pathway in human colorectal can. *Clinical and experimental pharmacology & physiology*, 45(7):729–741.

Howe, P. W., Nagai, K., Neuhaus, D., and Varani, G. (1994). NMR studies of U1 snRNA recognition by the N-terminal RNP domain of the human U1A protein. *The EMBO journal*, 13(16):3873–81.

Hua, Y., Sahashi, K., Rigo, F., Hung, G., Horev, G., Bennett, C. F., and Krainer, A. R. (2011). Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model. *Nature*, 478(7367):123–126.

Hulsen, T., de Vlieg, J., and Alkema, W. (2008). BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics*, 9(1):488.

Jangi, M., Boutz, P. L., Paul, P., and Sharp, P. A. (2014). Rbfox2 controls autoregulation in RNA-binding protein networks. *Genes & development*, 28(6):637–51.

Jangi, M. and Sharp, P. (2014). Building Robust Transcriptomes with Master Splicing Factors. *Cell*, 159(3):487–498.

Jin, D., Hidaka, K., Shirai, M., and Morisaki, T. (2010). RNA-binding motif protein 24 regulates myogenin expression and promotes myogenic differentiation. *Genes to Cells*, 15(11):1158–1167.

Kalsotra, A., Xiao, X., Ward, A. J., Castle, J. C., Johnson, J. M., Burge, C. B., and Cooper, T. A. (2008). A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proceedings of the National Academy of Sciences*, 105(51):20333–20338.

Kaufman, O. H., Lee, K., Martin, M., Rothhämel, S., and Marlow, F. L. (2018). rbpms2 functions in Balbiani body architecture and ovary fate. *PLOS Genetics*, 14(7):e1007489.

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., and Stamm, S. (2013). Function of alternative splicing. *Gene*, 514(1):1–30.

Khan, A. and Zhang, X. (2016). dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Research*, 44(D1):D164–D171.

Kino, Y., Washizu, C., Kurosawa, M., Oma, Y., Hattori, N., Ishiura, S., and Nukina, N. (2015). Nuclear localization of MBNL1: splicing-mediated autoregulation and repression of repeat-derived aberrant proteins. *Human Molecular Genetics*, 24(3):740–756.

Klemm, S. L., Shipony, Z., and Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics*, 20(4):207–220.

Kume, T., Jiang, H., Topczewska, J. M., and Hogan, B. L. (2001). The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. *Genes & Development*, 15(18):2470–2482.

Kurmangaliyev, Y. Z., Sutormin, R. A., Naumenko, S. A., Bazykin, G. A., and Gelfand, M. S. (2013). Functional implications of splicing polymorphisms in the human genome. *Human Molecular Genetics*, 22(17):3449–3459.

Kwong, J. M. K., Caprioli, J., and Piri, N. (2010). RNA binding protein with multiple splicing: a new marker for retinal ganglion cells. *Investigative ophthalmology & visual science*, 51(2):1052–8.

Kırlı, K., Karaca, S., Dehne, H. J., Samwer, M., Pan, K. T., Lenz, C., Urlaub, H., and Görlich, D. (2015). A deep proteomics perspective on CRM1-mediated nuclear export and nucleocytoplasmic partitioning. *eLife*, 4.

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., and Weirauch, M. T. (2018). The Human Transcription Factors. *Cell*, 172(4):650–665.

Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., and Brenner, S. E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, 446(7138):926–9.

Lee, T. and Young, R. (2013). Transcriptional Regulation and Its Misregulation in Disease. *Cell*, 152(6):1237–1251.

Li, B. and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1):323.

Li, J., Choi, P. S., Chaffer, C. L., Labella, K., Hwang, J. H., Giacomelli, A. O., Kim, J. W., Ilic, N., Doench, J. G., Ly, S. H., Dai, C., Hagel, K., Hong, A. L., Gjoerup, O., Goel, S., Ge, J. Y., Root, D. E., Zhao, J. J., Brooks, A. N., Weinberg, R. A., and Hahn, W. C. (2018). An alternative splicing switch in FLNB promotes the mesenchymal cell state in human breast cancer. *eLife*, 7.

Liang, X. and Howard, J. (2018). Structural Biology: Piezo Senses Tension through Curvature. *Current biology : CB*, 28(8):R357–R359.

Lin, Y., Protter, D. S. W., Rosen, M. K., and Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Molecular cell*, 60(2):208–19.

Llorian, M., Gooding, C., Bellora, N., Hallegger, M., Buckroyd, A., Wang, X., Rajgor, D., Kayikci, M., Feltham, J., Ule, J., Eyras, E., and Smith, C. W. J. (2016). The alternative splicing program of differentiated smooth muscle cells involves concerted non-productive splicing of post-transcriptional regulators. *Nucleic acids research*, 44(18):8933–8950.

Llorian, M., Schwartz, S., Clark, T. A., Hollander, D., Tan, L.-Y., Spellman, R., Gordon, A., Schweitzer, A. C., de la Grange, P., Ast, G., and Smith, C. W. J. (2010). Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nature Structural & Molecular Biology*, 17(99):1114–1123.

Llorian, M. and Smith, C. W. J. (2011). Decoding muscle alternative splicing. *Current opinion in genetics & development*, 21(4):380–387.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., Fleming, J., Siminoff, L., Traino, H., Mosavel, M., Barker, L., Jewell, S., Rohrer, D., Maxim, D., Filkins, D., Harbach, P., Cortadillo, E., Berghuis, B., Turner, L., Hudson, E., Feenstra, K., Sobin, L., Robb, J., Branton, P., Korzeniewski, G., Shive, C., Tabor, D., Qi, L., Groch, K., Nampally, S., Buia, S., Zimmerman, A., Smith, A., Burges, R., Robinson, K., Valentino, K., Bradbury, D., Cosentino, M., Diaz-Mayoral, N., Kennedy, M., Engel, T., Williams, P., Erickson, K., Ardlie, K., Winckler, W., Getz, G., DeLuca, D., MacArthur, D., Kellis, M., Thomson, A., Young, T., Gelfand, E., Donovan, M., Meng, Y., Grant, G., Mash, D., Marcus, Y., Basile, M., Liu, J., Zhu, J., Tu, Z., Cox, N. J., Nicolae, D. L., Gamazon, E. R., Im, H. K., Konkashbaev, A., Pritchard, J., Stevens, M., Flutre, T., Wen, X., Dermitzakis, E. T., Lappalainen, T., Guigo, R., Monlong, J., Sammeth, M., Koller, D., Battle, A., Mostafavi, S., McCarthy, M., Rivas, M., Maller, J., Rusyn, I., Nobel, A., Wright, F., Shabalin, A., Feolo, M., Sharopova, N., Sturcke, A., Paschal, J., Anderson, J. M., Wilder, E. L., Derr, L. K., Green, E. D., Struewing, J. P., Temple, G., Volpi, S., Boyer, J. T., Thomson, E. J., Guyer, M. S., Ng, C., Abdallah, A., Colantuoni, D., Insel, T. R., Koester, S. E., Little, A. R., Bender, P. K., Lehner, T., Yao, Y., Compton, C. C., Vaught, J. B., Sawyer, S., Lockhart, N. C., Demchok, J., and Moore, H. F. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, 45(6):580–585.

Lorson, C. L., Hahnen, E., Androphy, E. J., and Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proceedings of the National Academy of Sciences of the United States of America*, 96(11):6307–11.

Lovci, M. T., Bengtson, M. H., and Massirer, K. B. (2016). Post-Translational Modifications and RNA-Binding Proteins. In *Advances in experimental medicine and biology*, volume 907, pages 297–317.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550.

Lu, Y., Loh, Y.-H., Li, H., Cesana, M., Ficarro, S., Parikh, J., Salomonis, N., Toh, C.-X., Andreadis, S., Luckey, C., Collins, J., Daley, G., and Marto, J. (2014). Alternative Splicing of MBD2 Supports Self-Renewal in Human Pluripotent Stem Cells. *Cell Stem Cell*, 15(1):92–101.

Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell*, 144(1):16–26.

Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., and Misteli, T. (2010). Regulation of Alternative Splicing by Histone Modifications. *Science*, 327(5968):996–1000.

Lunde, B. M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nature Reviews Molecular Cell Biology*, 8(6):479–490.

Lundquist, E. A., Herman, R. K., Rogalski, T. M., Mullen, G. P., Moerman, D. G., and Shaw, J. E. (1996). The mec-8 gene of C. elegans encodes a protein with two RNA recognition motifs and regulates alternative splicing of unc-52 transcripts. *Development*, 122(5):1601–10.

Makeyev, E. V., Zhang, J., Carrasco, M. A., and Maniatis, T. (2007). The MicroRNA miR-124 Promotes Neuronal Differentiation by Triggering Brain-Specific Alternative Pre-mRNA Splicing. *Molecular Cell*, 27(3):435–448.

Maris, C., Dominguez, C., and Allain, F. H.-T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *The FEBS journal*, 272(9):2118–31.

Maslon, M. M., Braunschweig, U., Aitken, S., Mann, A. R., Kilanowski, F., Hunter, C. J., Blencowe, B. J., Kornblihtt, A. R., Adams, I. R., and Caceres, J. F. (2018). Slow transcriptional elongation causes embryonic lethality and perturbs kinetic coupling of long neural genes. *bioRxiv*, page 426577.

Masuda, A., Andersen, H. S., Doktor, T. K., Okamoto, T., Ito, M., Andresen, B. S., and Ohno, K. (2012). CUGBP1 and MBNL1 preferentially bind to 3' UTRs and facilitate mRNA decay. *Scientific reports*, 2(1):209.

Mayeda, A. and Krainer, A. R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell*, 68(2):365–75.

McGlincy, N. J. and Smith, C. W. J. (2008). Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends in biochemical sciences*, 33(8):385–93.

Meerbrey, K. L., Hu, G., Kessler, J. D., Roarty, K., Li, M. Z., Fang, J. E., Herschkowitz, J. I., Burrows, A. E., Ciccia, A., Sun, T., Schmitt, E. M., Bernardi, R. J., Fu, X., Bland, C. S., Cooper, T. A., Schiff, R., Rosen, J. M., Westbrook, T. F., and Elledge, S. J. (2011). The pINDUCER lentiviral toolkit for inducible RNA interference in

vitro and in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 108(9):3665–70.

Merkin, J., Russell, C., Chen, P., and Burge, C. B. (2012). Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. *Science*, 338(6114):1593–1599.

Michlewski, G. and Cáceres, J. F. (2019). Post-transcriptional control of miRNA biogenesis. *RNA*, 25(1):1–16.

Miyamoto, S., Hidaka, K., Jin, D., and Morisaki, T. (2009). RNA-binding proteins Rbm38 and Rbm24 regulate myogenic differentiation via p21-dependent and -independent regulatory pathways. *Genes to Cells*, 14(11):1241–1252.

Mockenhaupt, S. and Makeyev, E. V. (2015). Non-coding functions of alternative pre-mRNA splicing in development. *Seminars in Cell & Developmental Biology*, 47-48:32–39.

Monani, U. R., Lorson, C. L., Parsons, D. W., Prior, T. W., Androphy, E. J., Burghes, A. H., and McPherson, J. D. (1999). A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Human molecular genetics*, 8(7):1177–83.

Montes, M., Sanford, B. L., Comiskey, D. F., and Chandler, D. S. (2019). RNA Splicing and Disease: Animal Models to Therapies. *Trends in Genetics*, 35(1):68–87.

Muñoz, M. J., Santangelo, M. S. P., Paronetto, M. P., de la Mata, M., Pelisch, F., Boireau, S., Glover-Cutter, K., Ben-Dov, C., Blaustein, M., Lozano, J. J., Bird, G., Bentley, D., Bertrand, E., and Kornblihtt, A. R. (2009). DNA Damage Regulates Alternative Splicing through Inhibition of RNA Polymerase II Elongation. *Cell*, 137(4):708–720.

Naro, C. and Sette, C. (2013). Phosphorylation-mediated regulation of alternative splicing in cancer. *International journal of cell biology*, 2013:151839.

Neumann, A., Schindler, M., Olofsson, D., Wilhelmi, I., Schürmann, A., and Heyd, F. (2019). Genome-wide identification of alternative splicing events that regulate protein transport across the secretory pathway. *Journal of Cell Science*, Epub ahead:jcs.230201.

Nevins, J. R. and Darnell, J. E. (1978). Steps in the processing of Ad2 mRNA: Poly(A)+ Nuclear sequences are conserved and poly(A) addition precedes splicing. *Cell*, 15(4):1477–1493.

Ni, J. Z., Grate, L., Donohue, J. P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T. A., Blume, J. E., and Ares, M. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes & development*, 21(6):708–18.

Nickless, A., Bailis, J. M., and You, Z. (2017). Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell & bioscience*, 7:26.

Notarnicola, C., Rouleau, C., Le Guen, L., Virsolvy, A., Richard, S., Faure, S., and De Santa Barbara, P. (2012). The RNA-binding protein RBPMS2 regulates development of gastrointestinal smooth muscle. *Gastroenterology*, 143(3):687–97.e1–9.

Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H., and Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, 372(6505):432–438.

Owens, G. K., Kumar, M. S., and Wamhoff, B. R. (2004). Molecular Regulation of Vascular Smooth Muscle Cell Differentiation in Development and Disease. *Physiological Reviews*, 84(3):767–801.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics*, 40(12):1413–5.

Panettieri, R. A. (2002). Airway smooth muscle: an immunomodulatory cell. *The Journal of allergy and clinical immunology*, 110(6 Suppl):S269–74.

Park, E., Pan, Z., Zhang, Z., Lin, L., and Xing, Y. (2018). The Expanding Landscape of Alternative Splicing Variation in Human Populations. *The American Journal of Human Genetics*, 102(1):11–26.

Partridge, L. M. M. and Carter, D. A. (2017). Novel Rbfox2 isoforms associated with alternative exon usage in rat cortex and suprachiasmatic nucleus. *Scientific Reports*, 7(1):9929.

Pervouchine, D., Popov, I., Berry, A., Borsari, B., Frankish, A., and Guigo, R. (2018). Novel autoregulatory cases of alternative splicing coupled with nonsense-mediated mRNA decay. *bioRxiv*, page 464404.

Plaschka, C., Newman, A. J., and Nagai, K. (2019). Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast. *Cold Spring Harbor perspectives in biology*, Epub ahead:a032391.

Polymenidou, M. (2018). The RNA face of phase separation. *Science*, 360(6391):859–860.

Rambout, X., Detiffe, C., Bruyr, J., Mariavelle, E., Cherkaoui, M., Brohée, S., Demoitié, P., Lebrun, M., Soin, R., Lesage, B., Guedri, K., Beullens, M., Bollen, M., Farazi, T. A., Kettmann, R., Struman, I., Hill, D. E., Vidal, M., Kruys, V., Simonis, N., Twizere, J.-C., and Dequiedt, F. (2016). The transcription factor ERG recruits CCR4–NOT to control mRNA decay and mitotic progression. *Nature Structural & Molecular Biology*, 23(7):663–672.

Rastgoo, N., Pourabdollah, M., Abdi, J., Reece, D., and Chang, H. (2018). Dysregulation of EZH2/miR-138 axis contributes to drug resistance in multiple myeloma by downregulating RBPMS. *Leukemia*, 32(11):2471–2482.

Ray, D., Kazan, H., Chan, E. T., Castillo, L. P., Chaudhry, S., Talukder, S., Blencowe, B. J., Morris, Q., and Hughes, T. R. (2009). Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nature Biotechnology*, 27(7):667–670.

Reed, R. (1996). Initial splice-site recognition and pairing during pre-mRNA splicing. *Current opinion in genetics & development*, 6(2):215–20.

Retailleau, K., Duprat, F., Arhatte, M., Ranade, S. S., Peyronnet, R., Martins, J. R., Jodar, M., Moro, C., Offermanns, S., Feng, Y., Demolombe, S., Patel, A., and Honoré, E. (2015). Piezo1 in Smooth Muscle Cells Is Involved in Hypertension-Dependent Arterial Remodeling. *Cell reports*, 13(6):1161–1171.

Rheaume, B. A., Jereen, A., Bolisetty, M., Sajid, M. S., Yang, Y., Renna, K., Sun, L., Robson, P., and Trakhtenberg, E. F. (2018). Single cell transcriptome profiling of retinal ganglion cells identifies cellular subtypes. *Nature Communications*, 9(1):2759.

Roberts, G. C., Gooding, C., Mak, H. Y., Proudfoot, N. J., and Smith, C. W. (1998). Co-transcriptional commitment to alternative splice site selection. *Nucleic acids research*, 26(24):5568–72.

Roca, X., Krainer, A. R., and Eperon, I. C. (2013). Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes & development*, 27(2):129–44.

Rosenfeld, M. G., Lin, C. R., Amara, S. G., Stolarsky, L., Roos, B. A., Ong, E. S., and Evans, R. M. (1982). Calcitonin mRNA polymorphism: peptide switching associated with alternative RNA splicing events. *Proceedings of the National Academy of Sciences of the United States of America*, 79(6):1717–21.

Ross, R. (1993). The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature*, 362(6423):801–9.

Rothman, A., Kulik, T. J., Taubman, M. B., Berk, B. C., Smith, C. W., and Nadal-Ginard, B. (1992). Development and characterization of a cloned rat pulmonary arterial smooth muscle cell line that maintains differentiated properties through multiple subcultures. *Circulation*, 86(6):1977–86.

Roux, P. P. and Blenis, J. (2004). ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions. *Microbiology and molecular biology reviews : MMBR*, 68(2):320–44.

Sagnol, S., Marchal, S., Yang, Y., Allemand, F., and de Santa Barbara, P. (2016). Epithelial Splicing Regulatory Protein 1 (ESRP1) is a new regulator of stomach smooth muscle development and plasticity. *Developmental biology*, 414(2):207–18.

Sagnol, S., Yang, Y., Bessin, Y., Allemand, F., Hapkova, I., Notarnicola, C., Guichou, J.-F., Faure, S., Labesse, G., and de Santa Barbara, P. (2014). Homodimerization of RBPMS2 through a new RRM-interaction motif is necessary to control smooth muscle plasticity. *Nucleic acids research*, 42(15):10173–84.

Salz, H. K. (2011). Sex determination in insects: a binary decision based on alternative splicing. *Current Opinion in Genetics & Development*, 21(4):395–400.

Sammeth, M., Foissac, S., and Guigó, R. (2008). A General Definition and Nomenclature for Alternative Splicing Events. *PLoS Computational Biology*, 4(8):e1000147.

Sanders, K. M., Koh, S. D., Ro, S., and Ward, S. M. (2012). Regulation of gastrointestinal motility–insights from smooth muscle biology. *Nature reviews. Gastroenterology & hepatology*, 9(11):633–45.

Sato, K., Saigusa, D., Saito, R., Fujioka, A., Nakagawa, Y., Nishiguchi, K. M., Kokubun, T., Motoike, I. N., Maruyama, K., Omodaka, K., Shiga, Y., Uruno, A., Koshiba, S., Yamamoto, M., and Nakazawa, T. (2018). Metabolomic changes in the mouse retina after optic nerve injury. *Scientific Reports*, 8(1):11930.

Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., Dixon, J. E., and Zipursky, S. (2000). Drosophila Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity. *Cell*, 101(6):671–684.

Scotti, M. M. and Swanson, M. S. (2016). RNA mis-splicing in disease. *Nature Reviews Genetics*, 17(1):19–32.

Sharma, S., Maris, C., Allain, F.-T., and Black, D. (2011). U1 snRNA Directly Interacts with Polypyrimidine Tract-Binding Protein during Splicing Repression. *Molecular Cell*, 41(5):579–588.

Sharp, P. A. (2005). The discovery of split genes and RNA splicing. *Trends in biochemical sciences*, 30(6):279–81.

Shen, S., Park, J. W., Lu, Z.-x., Lin, L., Henry, M. D., Wu, Y. N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America*, 111(51):E5593–601.

Shimamoto, A., Kitao, S., Ichikawa, K., Suzuki, N., Yamabe, Y., Imamura, O., Tokutake, Y., Satoh, M., Matsumoto, T., Kuromitsu, J., Kataoka, H., Sugawara, K., Sugawara, M., Sugimoto, M., Goto, M., and Furuichi, Y. (1996). A unique human gene that spans over 230 kb in the human chromosome 8p11-12 and codes multiple family proteins sharing RNA-binding motifs. *Proceedings of the National Academy of Sciences of the United States of America*, 93(20):10913–7.

Shin, H. Y. (2018). Targeting Super-Enhancers for Disease Treatment and Diagnosis. *Molecules and cells*, 41(6):506–514.

Shinde, M. Y., Sidoli, S., Kulej, K., Mallory, M. J., Radens, C. M., Reicherter, A. L., Myers, R. L., Barash, Y., Lynch, K. W., Garcia, B. A., and Klein, P. S. (2017). Phosphoproteomics reveals that glycogen synthase kinase-3 phosphorylates multiple splicing factors and is associated with alternative splicing. *The Journal of biological chemistry*, 292(44):18240–18255.

Shukla, S. and Fisher, S. A. (2008). Tra2β As a Novel Mediator of Vascular Smooth Muscle Diversification. *Circulation Research*, 103(5):485–492.

Sims, R. J., Millhouse, S., Chen, C.-F., Lewis, B. A., Erdjument-Bromage, H., Tempst, P., Manley, J. L., and Reinberg, D. (2007). Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. *Molecular Cell*, 28(4):665–676.

Sinha, R., Kim, Y. J., Nomakuchi, T., Sahashi, K., Hua, Y., Rigo, F., Bennett, C. F., and Krainer, A. R. (2018). Antisense oligonucleotides correct the familial dysautonomia splicing defect in IKBKAP transgenic mice. *Nucleic Acids Research*, 46(10):4833–4844.

Skalska, L., Beltran-Nebot, M., Ule, J., and Jenner, R. G. (2017). Regulatory feedback from nascent RNA to chromatin and transcription. *Nature Reviews Molecular Cell Biology*, 18(5):331–337.

Soufari, H. and Mackereth, C. D. (2017). Conserved binding of GCAC motifs by MEC-8, couch potato, and the RBPMS protein family. *RNA*, 23(3):308–316.

Southby, J., Gooding, C., and Smith, C. W. (1999). Polypyrimidine tract binding protein functions as a repressor to regulate alternative splicing of alpha-actinin mutally exclusive exons. *Molecular and cellular biology*, 19(4):2699–711.

Spellman, R., Llorian, M., and Smith, C. W. J. (2007). Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Molecular cell*, 27(3):420–34.

Spellman, R. and Smith, C. W. (2006). Novel modes of splicing repression by PTB. *Trends in Biochemical Sciences*, 31(2):73–76.

Spin, J. M., Maegdefessel, L., and Tsao, P. S. (2012). Vascular smooth muscle cell phenotypic plasticity: focus on chromatin remodelling. *Cardiovascular Research*, 95(2):147–155.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T., and Soreq, H. (2005). Function of alternative splicing. *Gene*, 344:1–20.

Sun, Q. (2005). Defining the mammalian CArGome. *Genome Research*, 16(2):197–207.

Sun, S., Zhang, Z., Fregoso, O., and Krainer, A. R. (2012). Mechanisms of activation and repression by the alternative splicing factors RBFOX1/2. *RNA*, 18(2):274–283.

Sun, Y., Ding, L., Zhang, H., Han, J., Yang, X., Yan, J., Zhu, Y., Li, J., Song, H., and Ye, Q. (2006). Potentiation of Smad-mediated transcriptional activation by the RNA-binding protein RBPMS. *Nucleic Acids Research*, 34(21):6314–6326.

Szczot, M., Pogorzala, L. A., Solinski, H. J., Young, L., Yee, P., Le Pichon, C. E., Chesler, A. T., and Hoon, M. A. (2017). Cell-Type-Specific Splicing of Piezo2 Regulates Mechanotransduction. *Cell Reports*, 21(10):2760–2771.

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Roth, A., Bork, P., Jensen, L. J., and von Mering, C. (2017). The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research*, 45(D1):D362–D368.

Sznajder, Ł. J., Michalak, M., Taylor, K., Cywoniuk, P., Kabza, M., Wojtkowiak-Szlachcic, A., Matłoka, M., Konieczny, P., and Sobczak, K. (2016). Mechanistic determinants of MBNL activity. *Nucleic acids research*, 44(21):10326–10342.

Tabaglio, T., Low, D. H., Teo, W. K. L., Goy, P. A., Cywoniuk, P., Wollmann, H., Ho, J., Tan, D., Aw, J., Pavesi, A., Sobczak, K., Wee, D. K. B., and Guccione, E. (2018). MBNL1 alternative splicing isoforms play opposing roles in cancer. *Life science alliance*, 1(5):e201800157.

Teplova, M., Farazi, T. A., Tuschl, T., and Patel, D. J. (2016). Structural basis underlying CAC RNA recognition by the RRM domain of dimeric RNA-binding protein RBPMS. *Quarterly reviews of biophysics*, 49:e1.

Thyberg, J. (1996). Differentiated properties and proliferation of arterial smooth muscle cells in culture. *International review of cytology*, 169:183–265.

Tran, H., Gourrier, N., Lemercier-Neuillet, C., Dhaenens, C.-M., Vautrin, A., Fernandez-Gomez, F. J., Arandel, L., Carpentier, C., Obriot, H., Eddarkaoui, S., Delattre, L., Van Brussels, E., Holt, I., Morris, G. E., Sablonnière, B., Buée, L., Charlet-Berguerand, N., Schraen-Maschke, S., Furling, D., Behm-Ansmant, I., Branlant, C., Caillet-Boudin, M.-L., and Sergeant, N. (2011). Analysis of exonic regions involved in nuclear localization, splicing activity, and dimerization of Muscleblind-like-1 isoforms. *The Journal of biological chemistry*, 286(18):16435–46.

van der Flier, A., Kuikman, I., Kramer, D., Geerts, D., Kreft, M., Takafuta, T., Shapiro, S. S., and Sonnenberg, A. (2002). Different splice variants of filamin-B affect myogenesis, subcellular distribution, and determine binding to integrin β subunits. *The Journal of cell biology*, 156(2):361–76.

van der Veer, E. P., de Bruin, R. G., Kraaijeveld, A. O., de Vries, M. R., Bot, I., Pera, T., Segers, F. M., Trompet, S., van Gils, J. M., Roeten, M. K., Beckers, C. M., van Santbrink, P. J., Janssen, A., van Solingen, C., Swildens, J., de Boer, H. C., Peters, E. a., Bijkerk, R., Rousch, M., Doop, M., Kuiper, J., Schalij, M. J., van der Wal, A. C., Richard, S., van Berkel, T. J. C., Pickering, J. G., Hiemstra, P. S., Goumans, M. J., Rabelink, T. J., de Vries, A. a. F., Quax, P. H. a., Jukema, J. W., Biessen, E. a. L., and van Zonneveld, A. J. (2013). Quaking, an RNA-binding protein, is a critical regulator of vascular smooth muscle cell phenotype. *Circulation research*, 113(9):1065–75.

Venables, J. P., Vignal, E., Baghdiguian, S., Fort, P., and Tazi, J. (2012). Tissue-Specific Alternative Splicing of Tak1 Is Conserved in Deuterostomes. *Molecular Biology and Evolution*, 29(1):261–269.

Venigalla, R. K. C. and Turner, M. (2012). RNA-binding proteins as a point of convergence of the PI3K and p38 MAPK pathways. *Frontiers in Immunology*, 3:398.

Vu, N. T., Park, M. A., Shultz, J. C., Goehe, R. W., Hoeferlin, L. A., Shultz, M. D., Smith, S. A., Lynch, K. W., and Chalfant, C. E. (2013). hnRNP U Enhances Caspase-9 Splicing and Is Modulated by AKT-dependent Phosphorylation of hnRNP L. *Journal of Biological Chemistry*, 288(12):8575–8584.

Waites, G. T., Graham, I. R., Jackson, P., Millake, D. B., Patel, B., Blanchard, A. D., Weller, P. A., Eperon, I. C., and Critchley, D. R. (1992). Mutually exclusive splicing of calcium-binding domain exons in chick alpha-actinin. *The Journal of biological chemistry*, 267(9):6263–71.

Wang, D. and Atanasov, A. G. (2019). The microRNAs Regulating Vascular Smooth Muscle Cell Proliferation: A Minireview. *International Journal of Molecular Sciences*, 20(2):324.

Wang, E. T., Cody, N. A. L., Jog, S., Biancolella, M., Wang, T. T., Treacy, D. J., Luo, S., Schroth, G. P., Housman, D. E., Reddy, S., Lécuyer, E., and Burge, C. B. (2012). Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell*, 150(4):710–24.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., and Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221):470–6.

Wehrle-Haller, B. (2012). Structure and function of focal adhesions. *Current Opinion in Cell Biology*, 24(1):116–124.

Weischenfeldt, J., Waage, J., Tian, G., Zhao, J., Damgaard, I., Jakobsen, J., Kristiansen, K., Krogh, A., Wang, J., and Porse, B. T. (2012). Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. *Genome Biology*, 13(5):R35.

White, E. S., Baralle, F. E., and Muro, A. F. (2008). New insights into form and function of fibronectin splice variants. *The Journal of pathology*, 216(1):1–14.

Wieczorek, D. F., Smith, C. W., and Nadal-Ginard, B. (1988). The rat alpha-tropomyosin gene generates a minimum of six different mRNAs coding for striated, smooth, and nonmuscle isoforms by alternative splicing. *Molecular and cellular biology*, 8(2):679–94.

Witten, J. T. and Ule, J. (2011). Understanding splicing regulation through RNA splicing maps. *Trends in genetics : TIG*, 27(3):89–97.

Wollerton, M. C., Gooding, C., Robinson, F., Brown, E. C., Jackson, R. J., and Smith, C. W. (2001). Differential alternative splicing activity of isoforms of polypyrimidine tract binding protein (PTB). *RNA*, 7(6):819–32.

Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. a., and Smith, C. W. J. (2004). Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Molecular cell*, 13(1):91–100.

Wongpalee, S. P., Vashisht, A., Sharma, S., Chui, D., Wohlschlegel, J. A., and Black, D. L. (2016). Large-scale remodeling of a repressed exon ribonucleoprotein to an exon definition complex active for splicing. *eLife*, 5.

Worth, N. F., Rolfe, B. E., Song, J., and Campbell, G. R. (2001). Vascular smooth muscle cell phenotypic modulation in culture is associated with reorganisation of contractile and cytoskeletal proteins. *Cell motility and the cytoskeleton*, 49(3):130–45.

Xie, N., Chen, M., Dai, R., Zhang, Y., Zhao, H., Song, Z., Zhang, L., Li, Z., Feng, Y., Gao, H., Wang, L., Zhang, T., Xiao, R.-P., Wu, J., and Cao, C.-M. (2017). SRSF1 promotes vascular smooth muscle cell proliferation through a Δ133p53/EGR1/KLF5 pathway. *Nature Communications*, 8:16016.

Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2017). Structure of a yeast step II catalytically activated spliceosome. *Science*, 355(6321):149–155.

Yap, K., Mukhina, S., Zhang, G., Tan, J. S., Ong, H. S., and Makeyev, E. V. (2018). A Short Tandem Repeat-Enriched RNA Assembles a Nuclear Compartment to Control Alternative Splicing and Promote Cell Survival. *Molecular Cell*, 72(3):525–540.e13.

Ye, G. J., Nesmith, A. P., and Parker, K. K. (2014). The Role of Mechanotransduction on Vascular Smooth Muscle Myocytes Cytoskeleton and Contractile Function. *The Anatomical Record*, 297(9):1758–1769.

Ye, L., Gu, L., Caprioli, J., and Piri, N. (2018). RNA-binding protein Rbpms is represented in human retinas by isoforms A and C and its transcriptional regulation involves Sp1-binding site. *Molecular genetics and genomics : MGG*, 293(4):819–830.

Yee, B. A., Pratt, G. A., Graveley, B. R., Van Nostrand, E. L., and Yeo, G. W. (2019). RBP-Maps enables robust generation of splicing regulatory maps. *RNA*, 25(2):193–204.

Ying, Y., Wang, X.-J., Vuong, C. K., Lin, C.-H., Damianov, A., and Black, D. L. (2017). Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. *Cell*, 170(2):312–323.e10.

Young, R. (2011). Control of the Embryonic Stem Cell State. *Cell*, 144(6):940–954.

Zhang, C., Zhang, Z., Castle, J., Sun, S., Johnson, J., Krainer, A. R., and Zhang, M. Q. (2008). Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes & Development*, 22(18):2550–2563.

Zhang, X., Yan, C., Hang, J., Finci, L. I., Lei, J., and Shi, Y. (2017). An Atomic Structure of the Human Spliceosome. *Cell*, 169(5):918–929.e14.

Zheng, C. L., Fu, X.-D., and Gribskov, M. (2005). Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. *RNA*, 11(12):1777–87.

Chapter

Appendix

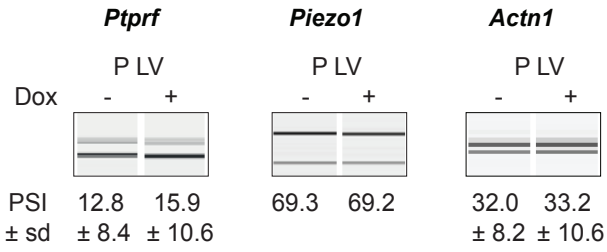


Fig. A.1 Doxycycline treatment of lentiviral control cell lines does not affect AS. Validation of RBPMS regulated ASE in lentiviral control populations (P LV) upon doxycycline treatment. A differentiated and proliferative cassette exon, *Ptprf* and *Piezo1* respectively, and a mutually exclusive exon *Actn1* are shown. Values shown are the mean ± sd of PSI (n = 3) of percentage of exon inclusion (PSI). Statistical significance was calculated using Student's t-test (* $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$). Experiment was carried out by Clare Gooding.

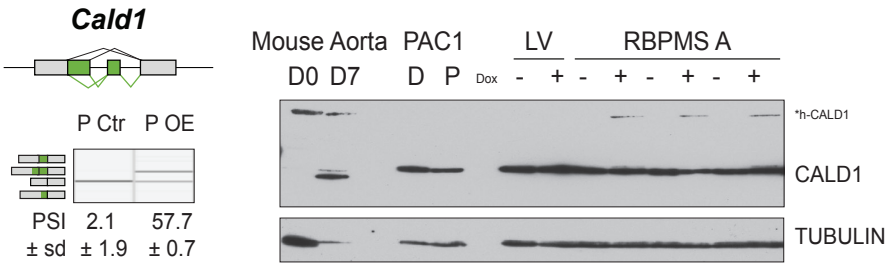


Fig. A.2 RBPMS overexpression promotes smooth muscle CALD1 isoform switch. Left, RT-PCR validation of *Cald1* ASE (A5SS and SE) in the RBPMS overexpression in proliferative PAC1 cells. Schematic of the *Cald1* ASE and *Cald1* isoforms amplified in the PCR are shown on the top and left, respectively. **Right**, the switch in *Cald1* isoform was shown at the protein level by western blot probing for CALD1. The specificity of the isoform switch was also confirmed in *Mus musculus* differentiated (D0) and proliferative (D7) SMC samples from aorta tissue. The larger smooth muscle specific isoform of CALD1 is indicated by h-CALD1. Values shown are mean ± sd (n = 3) of percentage of exon inclusion. Statistical significance was calculated using Student's t-test and is shown as * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Experiment was carried out by Clare Gooding.

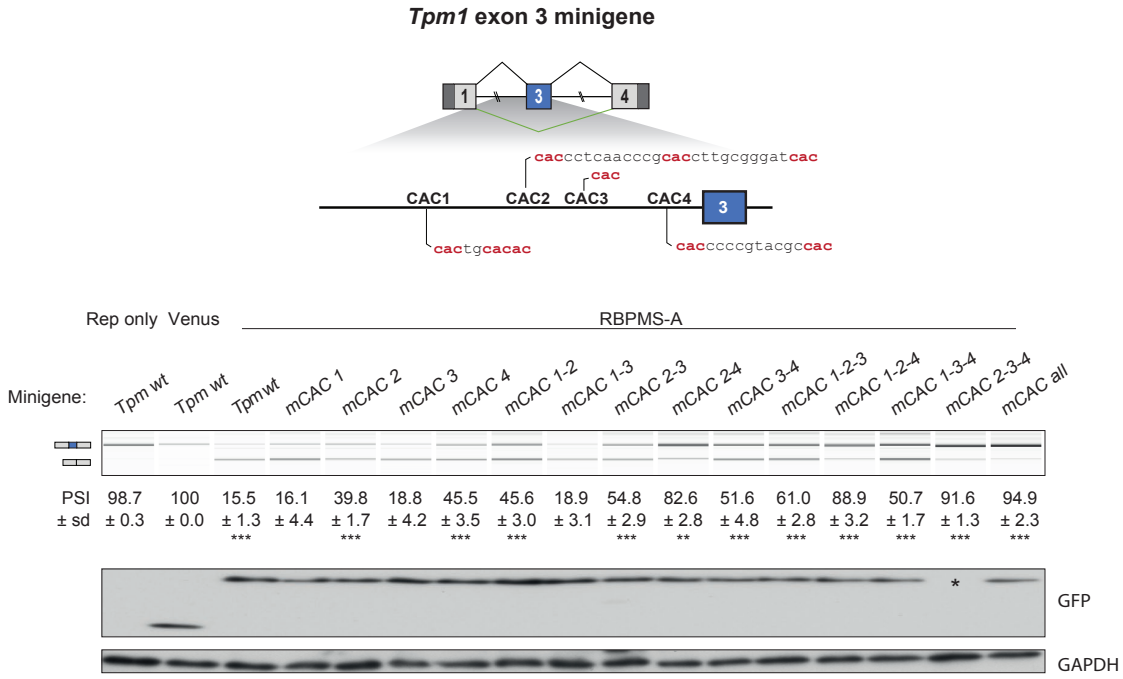


Fig. A.3 Mapping RBPMS cis elements in the *Tpm1* exon 3 splicing reporter. Top, potential RBPMS CAC motifs upstream of exon 3 are highlighted in the schematic. **Middle**, combinations of different CAC mutations were tested for RBPMS A overexpression in HEK293 cells. Schematics of the splicing isoforms identify the PCR products and values shown are mean ± sd (n = 3) of PSI. In case of MXE the Sm exon inclusion is shown. Statistical significance was calculated using Student's t-test (* $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$). **Bottom**, overexpression was assessed by western blot probing for GFP and GAPDH as a loading control. * No overexpression was detected in this lane although the PSI differs from the reporter only PSI. Experiment was carried out by Fredderick Richards.

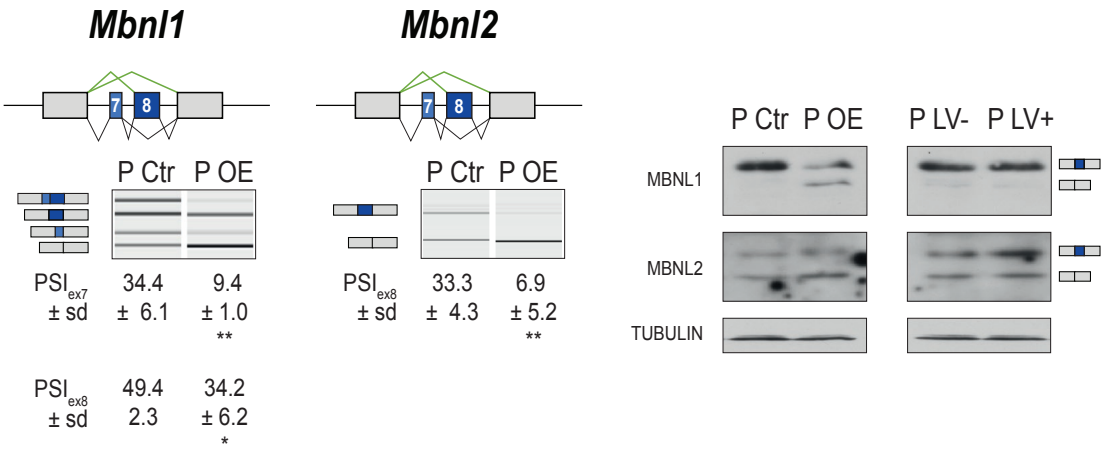


Fig. A.4 RBPMS overexpression switches MBNL1 and 2 splicing isoforms. **Left**, RT-PCR validation of *Mbnl1* exons 7 and 8 and *Mbnl2*. Values shown are the mean of the PSI ± sd (n = 3). PSI values are shown for each *Mbnl1* exon (exon 7 – 36 bp and exon 8 – 95 bp) as PSI_{ex7} and PSI_{ex8}. For *Mbnl2*, exon 7 isoforms were not detected in the RT-PCRs. Schematics of the splicing isoforms identify the PCR products. **Right**, western blots probing for MBNL1 and 2 with splicing isoforms indicated on the right. TUBULIN was used as a loading control for the western blots. Statistical significance was calculated using Student's t-test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Experiment was carried out by Clare Gooding.

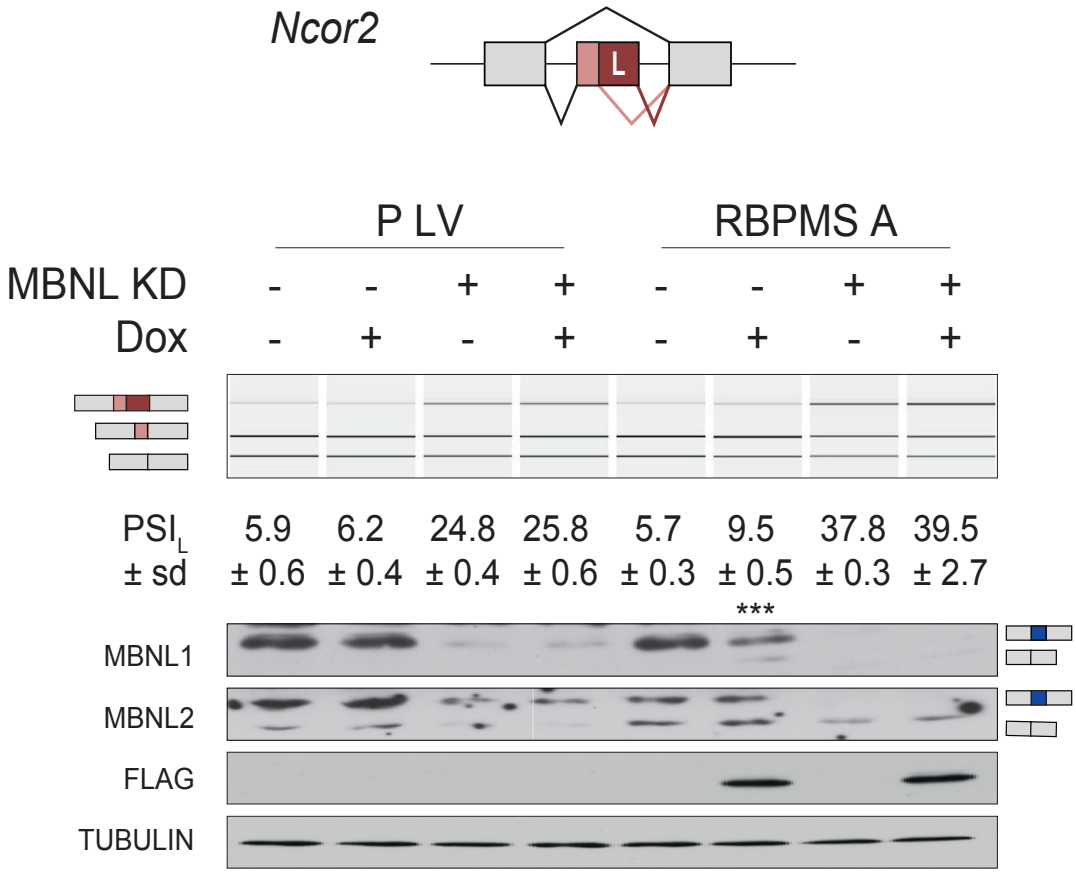


Fig. A.5 *Ncor2* is a RBPMS indirect splicing target regulated via MBNL isoform switch. Top, MBNL1 and 2 knockdown in inducible RBPMS A (P OE) or in lentiviral control (P LV) PAC1 cells to assess the dependence of the MBNL1 isoform switch in the regulation of *Ncor2* A5SS by RBPMS. *Ncor2* A5SS splicing was verified by RT-PCRs and the mean of the PSI ± sd (n = 3) are shown. PCR products are indicated by the splicing schematics on the side. **Bottom**, western blots for MBNL1 and 2 and FLAG to confirm MBNL knockdown and RBPMS A overexpression. TUBULIN was used as a loading control for the western blots. Statistical significance from Student's t-test is shown as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Experiment was carried out by Clare Gooding.

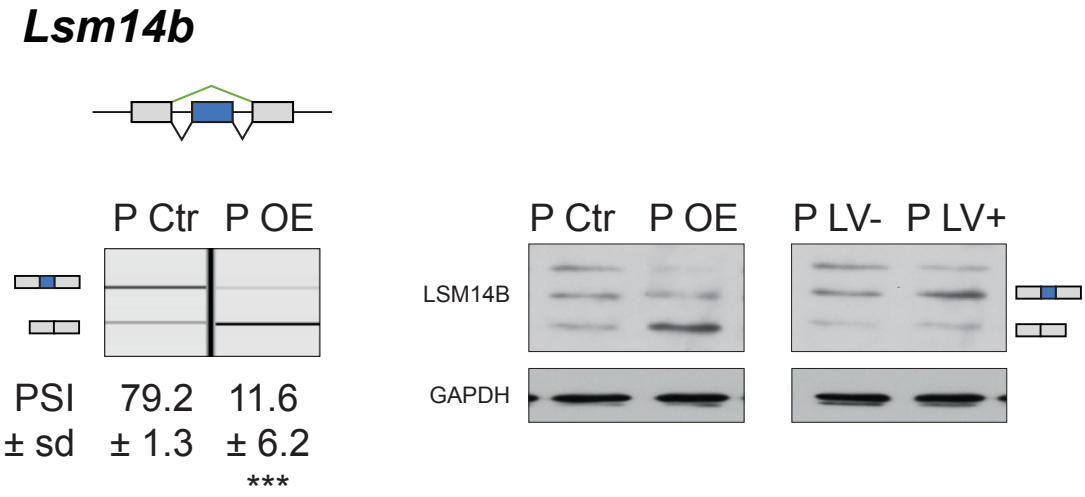


Fig. A.6 RT-PCR validation of *Lsm14b* in the RBPMS-A overexpression in proliferative PAC1 cells. **Left**, schematic of the *Lsm14b* event at the top and schematic of its isoforms at the left. **Right**, western blots probing for LSM14B with splicing isoforms indicated on the right. GAPDH was used as a loading control. Statistical significance was calculated using Student's t-test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Experiment was conducted by Clare Gooding.

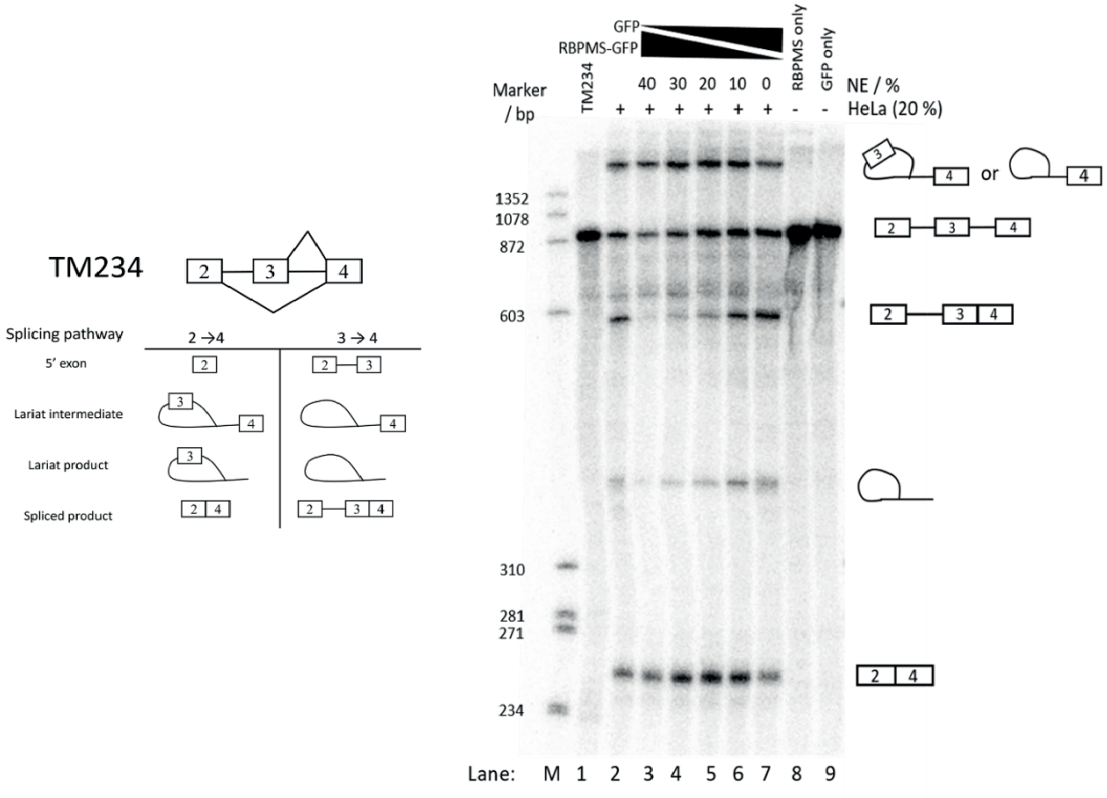


Fig. A.7 RBPMS represses the splicing of *Tpm1* exon 3 in vitro. Left, schematic representation of transcript TM234. Diagonal lines indicate alternative 2-4 or 3-4 splice pathways- expected intermediates and products of these pathways are shown below. **Right**, denaturing gel resolving radiolabelled products following in vitro splicing of TM234. Each reaction took place in a total volume of 10 μL. Nuclear extract (NE) made up 60% of the total reactions. Lanes 3- 7 contain 20% HeLa NE (final concentration 1.25 mg/mL) and a 40% cross-titration of NE from HEK293 cells overexpressing GFP against those overexpressing RBPMS-GFP (final concentration 0. 45 mg/mL). The HeLa only reaction (lane 2) is made up with Buffer E. Lanes 8 and 9 contains 40% NE as indicated plus 20% buffer E. The identities of various products are shown schematically to the right of the gel. Lariat products resolving above the pre-mRNA cannot be distinguished without debranching and subsequent analysis of their molecular weights. Experiment was conducted by Fredderick Richards and the figure and respective legend is from his dissertation.

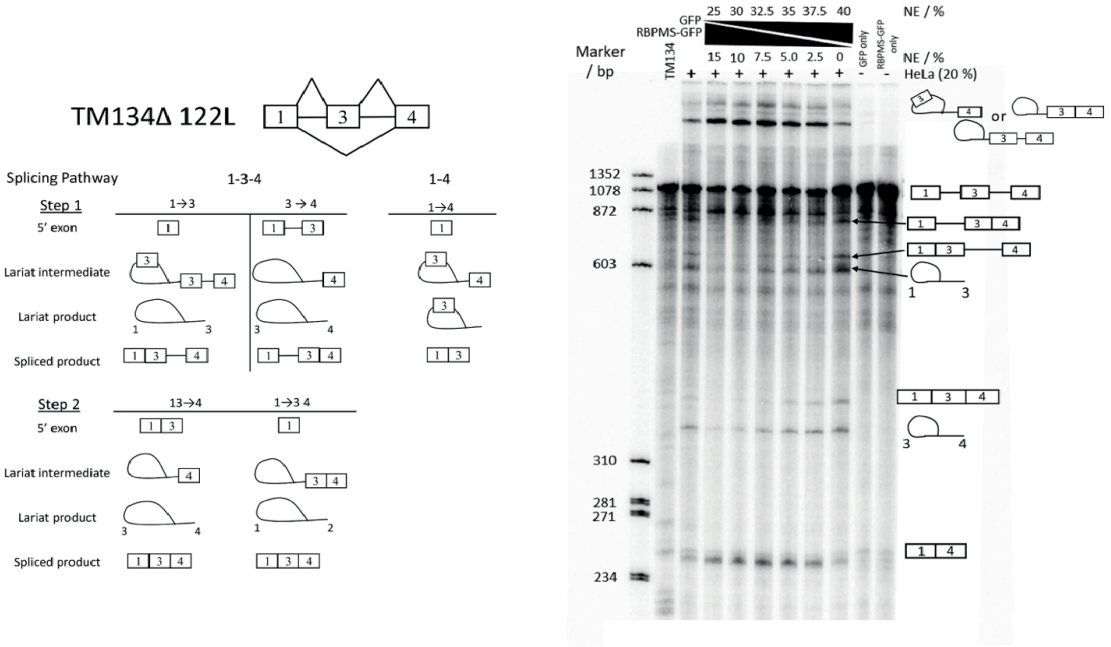


Fig. A.8 RBPMS represses the splicing of *Tpm1* exon 3 in vitro. Left, schematic representation of transcript TM134Δ122L. Diagonal lines indicate alternative 1-3-4 or 1-4 splice pathways- expected intermediates and products of these pathways are shown below. The precise- 1-3-4 splicing pathway depends on whether the upstream intron (1-3) or downstream intron (3-4) is removed first. **Right** In vitro splicing of TM134Δ122L Each reaction took place in a total volume of 10 µl. Nuclear extract (NE) made up 60% of the total reaction. Lanes 3- 8 contain 20% HeLa NE (final concentration 1.25 mg/ml) with titration of RBPMS-GFP NE titration from 15% to 0% as shown above each lane. This was titrated against GFP NE to a total volume of 40% to retain an overall protein content equal to the TM234 splicing reactions (i.e. 0.45 mg/mL final). The HeLa only reaction (lane 2) was made up with buffer E. Lanes 9 and 10 contain 40% NE as indicated plus 20% Buffer E. The identities of various products are shown schematically to the right of the gel. Lariat products resolving above the pre-mRNA cannot be distinguished without debranching and subsequent analysis of their molecular weights. Experiment was conducted by Fredderick Richards and the figure and respective legend is from his dissertation.

Table A1 Gene list of GO terms enriched in genes differentially spliced by RBPMS knockdown

GO Term	Description	Genes
Cell Component		
GO:0005856	Cytoskeleton	*SPATA7, LDB3, MYO9B , PDLIM5, ABI1, LRRFIP1, MPRIP, PDLIM7, OSBPL10, MAST2, FUZ, PARD3, PACSIN3, SYNE2, SGCA, FLNA, FLNB, ITGB1BP1, NF2, SVIL, NSMF, MACF1, RAI14, SMTN*
GO:0042995	Cell projection	*PKD1, PDLIM5, ACTN1, ABI1, SLC9A1, DAAM1, LRRC16, CAMK2G, FUZ, ARHGEF6, MARK3, FLNA, ITGB1BP1, SEMA6A, NF2, SVIL, NSMF, LDB3, CAD, CDC42BP, DNM2, PDLIM7, PIAS3, PAFAH1B1, DNM1, APP, SLC4A7, EIF5A, BBS2, FAT1, PTPRF*
GO:0030054	Cell junction	*CD99, PDLIM5, ACTN1, ABI, MPRIP, SLC9A1, FLNA, FLNB, PPFIA1, CLSTN1, NF2, SVIL, PPFIBP1, NSMF, MAP4K4, ARHGAP17, CDC42BPA, DNM2, PDLIM7, SORBS1, PARD3, SGCA, SYNE2, APP, ECT2, GJC1, FAT1, RAI1*
GO:0120025	Plasma membrane bounded cell projection	*PKD1, PDLIM5, ACTN1, ABI1, SLC9A1, DAAM1, LRRC16A, CAMK2G, ARHGEF6, MARK3, FLNA, ITGB1BP1, SEMA6A, NF2, SVIL, NSMF, LDB3, CDC42BPA, DNM2, PDLIM7, PIAS3, PAFAH1B1, DNM1, APP, SLC4A7, EIF5A, BBS2, FAT1, PTPRF*
GO:0005911	Cell-cell junction	*ACTN1, PDLIM5, ARHGAP17, CDC42BPA, PDLIM7, SLC9A1, SORBS1, PARD3, SGCA, FLNA, ECT2, GJC1, FAT1*
Molecular Function		
GO:0008092	Cytoskeletal protein binding	*ACTN1, PDLIM5, ABI1, MPRIP, DAAM1, FLNA, FLNB, CLSTN1, INF2, NF2, SVIL, MACF1, MAP4K4, LDB3, MYO9B, DNM2, MAST2, SORBS1, PACSIN3, FMNL3, PAFAH1B1, KLC1, STIM1, SYNE2, DNM1, RAB11FIP5, SMTN*
GO:0003779	Actin binding	*MYO9B, ACTN1, PDLIM5, MPRIP, DAAM1, SORBS1, FMNL3, SYNE2, FLNA, FLNB, INF2, NF2, SVIL, MACF1, SMTN*
GO:0030695	GTPase regulator activity	*ARHGEF6, MYO9B, ARHGAP23, GAPVD1, ARHGAP17, TBC1D1, ITGB1BP1, ARHGEF10L, AGFG2, EVI5L, SLIT2, ECT2*
GO:0017016	Ras GTPase binding	*MYO9B, GAPVD1, ARHGAP17, ARHGEF10L, DAAM1, EVI5L, CAMK2Ga, ARHGEF6, FMNL3, TBC1D1, FLNA, SBF1, RAB11FIP5, INF2, ECT2, ERBB2*
GO:0060589	Nucleoside-triphosphatase regulator activity	*MYO9B, GAPVD1, ARHGAP17, ARHGEF10L, EVI5L, SLIT2, ARHGEF6, ARHGAP23, FNIP2, TBC1D1, ITGB1BP1, AGFG2, ECT2*
Biological Process		
GO:0030029	Actin filament-based process	*LDB3, MYO9B, CDC42BPA, ARHGEF10L, SLC9A1, PDLIM7, DAAM1, SORBS1, FMNL3, PAFAH1B1, SYNE2, FLNA, MYL6, FLNB, INF2, NF2, SMTN*
GO:0044087	Regulation of cellular component biogenesis	*PDLIM5, ARHGEF10L, SLC9A1, SLIT2, LRRC16A, FUZ, PHLDB1, MTMR3, FLNA, ITGB1BP1, PPFIA1, CLSTN1, NF2, MACF1, CEP120, PIEZO1, MAP4K4, DNM2, STAG2, EVI5L, SDCCAG3, FNIP2, SYNE2, APP, PLD1, ECT2, MED25*
GO:0051128	Regulation of cellular component organization	*PKD1, CBFA2T2, ITGA7, PDLIM5, FN1, ARHGEF10L, SLC9A1, SLIT2, LRRC16A, FUZ, PHLDB1, MTMR3, FLNA, ITGB1BP1, PPFIA1, CLSTN1, SEMA6A, LGALS1, INF2, NF2, MAP2K7, NSMF, PIEZO1, CEP120, MACF1, MAP4K4, WDR70, ARHGAP17, DNM2, STAG2, EVI5L, PACSIN3, FMNL3, SDCCAG3, RHOT1, PAFAH1B1, SYNE2, FNIP2, DNM1, APP, EIF5A, PLD1, ECT2, MED25, ERBB2, PTPRF, CISH*
GO:0010810	Regulation of cell-substrate adhesion	*FLNA, ITGB1BP1, FN1, DNM2, ABI3BP, NF2, SLC9A1, LRRC16A, MACF1, MAP4K4*
GO:0007010	Cytoskeleton organization	*LDB3, CDC42BPA, ARHGEF10L, PDLIM7, DAAM1, MAST2, SORBS1, PARD3, PACSIN3, FMNL3, MARK3, PAFAH1B1, SYNE2, FLNA, FLNB, SEMA6A, INF2, NF2, SVIL, MACF1, CEP120, SMTN*